上海市“085工程”资助出版精品教材

预测与决策
——理论及应用

张华歆 主编

本书上篇介绍预测理论及方法，下篇介绍决策理论及方法；
介绍预测与决策的各种模型，并通过案例进行分析；
所选取的案例多以海运、港口问题为背景，适合高等院校经济与
管理专业、交通运输及航运物流专业使用。

上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

巍巍交大 百年书香
www.jiaodapress.com.cn
bookinfo@sjtu.edu.cn

责任编辑 冯 颖
封面设计 孙 敏
责任营销 韩长亮

上海市“085工程”资助出版精品教材

预测与决策——理论及应用

管理信息系统案例集萃

船舶CAD/CAM

海洋工程结构设计

流体机械泵与风机

船舶结构有限元建模与分析

交通工程专业英语

船舶电子电气英语

扫描二维码
关注上海交通大学出版社
“书香交大”

ISBN 978-7-313-10751-0

定价：42.00元

上架建议：交通运输

上海市“085工程”资助出版精品教材

预测与决策

——理论及应用

张华歆　主编

上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

内容提要

本书共15章，分为两大篇。上篇是预测理论与方法，共8章，系统介绍各种经济预测方法，包括：定性预测法、线性回归分析、时间序列分析、趋势曲线模型预测法、灰色预测等；下篇是决策理论与方法，共7章，系统介绍了确定型、非确定型和风险型决策的方法，同时也介绍多目标决策等内容。

本书中选取的例题多以海运、港口管理问题为背景，适合高等院校经济与管理专业、交通运输及航运物流等专业的本科生和硕士生使用，也适合作为政府、企事业管理干部、工程技术人员和理工科学生自学现代管理预测和决策方法的参考书。

图书在版编目(CIP)数据

预测与决策：理论与应用 / 张华歆主编．—上海：
上海交通大学出版社，2014
ISBN 978-7-313-10751-0

Ⅰ．①预…　Ⅱ．①张…　Ⅲ．①决策预测　Ⅳ．
①C934

中国版本图书馆CIP数据核字(2013)第312270号

预测与决策
——理论及应用

主　　编：张华歆
出版发行：上海交通大学出版社　　地　　址：上海市番禺路951号
邮政编码：200030　　电　　话：021-64071208
出 版 人：韩建民
印　　制：常熟市梅李印刷有限公司　　经　　销：全国新华书店
开　　本：787 mm×960 mm　1/16　　印　　张：22.75
字　　数：426千字
版　　次：2014年6月第1版　　印　　次：2014年6月第1次印刷
书　　号：ISBN 978-7-313-10751-0/C
定　　价：42.00元

前　言

在现代管理学中，预测学是关于预测理论与方法的科学体系，它突破了自然科学与社会科学两大科学的界限，是制订科学决策和科学计划的前提；决策是人们为了实现特定的目标，在占有大量调研预测资料的基础上，运用科学的理论和方法，充分发挥人的智慧，系统地分析主客观条件，围绕既定目标拟订各种实施预选方案，从若干个有价值的目标方案中选择或实施一个最佳执行方案的活动。

随着我国市场经济体制的逐步完善及我国交通运输对外开放领域的不断拓宽，世界主要航运公司纷纷打入我国航运市场，竞争日趋激烈。培养和造就一大批合格的适应各种复杂环境的高质量的交通运输、航运、港口等方面的管理人才是我国航运企业向高等教育提出的客观要求，是我们交通运输学科教师的重要任务，而如何将现代预测与决策基础理论恰当地应用至交通、航运、港口管理等领域，是上述任务的重要内容之一。鉴于此，本书具有以下特点：

一、对现有预测和决策方法进行了有效的归纳。本书把预测方法大致归纳成定性预测法、回归预测法和时间序列预测法三类，把决策方法归纳成确定型决策方法、不确定型决策方法、风险型决策方法和多目标决策方法。

二、充分体现预测和决策基础理论在交通运输、航运港口领域的应用。本书内的例题大多数都以相应航运、运输领域的管理问题为背景，以便于使交通运输、航运、港口管理专业的学生就预测和决策技术在实际管理工作中的应用有更深入的理解。

三、强调计算机在“预测和决策”中的运用。为了使“预测和决策”这门课程适应现代化发展的需要，我们在有关教学内容中使用了 Excel 等软件，使部分原本无法用手工完成的计算工作成为一件轻而易举的事情。这是本教材的第三个特点。

本书由上海海事大学交通运输学院硕士生导师张华歆副教授负责全书整体框架的设计及其全书的修改、总纂和定稿。在资料收集和整理方面，预测理论与方法部分，由王书琪负责完成；决策理论与方法部分，由张丽杰负责完成。本书的出版

得到了“上海市‘085 工程’资助出版精品教材”资助，并得到上海海事大学校基金项目“信息通信技术(ICT)作用下基于出行链的出行行为决策模型(20120081)”的资助。

“预测与决策”在我国尚属一门新学科，内容非常丰富，无论在理论上还是在实践上，都需要做深入的探讨。限于水平，本书难免存在不尽如人意之处，恳切希望广大教师和同学提出宝贵意见，以使其日臻完善。

张华歆

2014 年 1 月

目　　录

上篇　预测理论与方法

下篇　决策理论与方法

上篇　预测理论与方法

1　预测概述

1.1　预测的基本知识

预测学是关于预测理论与方法的科学体系，它突破了自然科学与社会科学两大科学的界限。科学预测是制订科学决策和科学计划的前提。本章将在介绍预测的概念、特点和作用的基础上，详细阐述预测的分类、步骤、误差及精度分析。

1.1.1　预测的定义

关于预测的定义，不同学者有不同观点。概括来讲，主要有以下几种：

(1) 预测，顾名思义就是一种预计和推测，可以预计事物发展的趋势，带有定性性质；也可以推测事物发展的指标，带有定量性质。

(2) 预测是人们事先对某一事物发展趋势和水平的一种测算。测算是指用数词和量词表示事物上升、下降、波动的规则或不规则趋势。

(3) 预测就是根据以往推测未来，根据已知推测未知，或以过去为基础推测未来，以昨天为依据估算今天。

(4) 预测就是对未来发展发表准确的意见。

(5) 预测就是预见、预言。

(6) 预测是一门正在形成的、研究未来和创造未来的科学。

(7) 预测从广义上讲，是指对未出现或出现了但还未被人们得知的事情作出推断和测算。

(8) 预测就是对事物未来作出科学的估计。“估”是指预测者凭借经验，对事物过去和现在的理解和认识；“计”包含计算、判断、推测等含义。

综合上述观点，本书给出预测的定义如下：所谓预测，就是人们根据事物以往发展的客观规律性和当前出现的各种可能性，运用科学的知识、方法和手段，对事物未来发展趋势和状态预先作出科学的估计和评价。该定义包括以下几个要点：

(1) 事物过去发展变动的过程和某些规律性是预测的根据。

(2) 预测时要参照事物当前正在出现的或已经出现的各种可能性。

(3) 预测要运用科学的方法和手段，如现代管理学、数学、统计学等方法。

(4) 预测的目的是分析事物的发展变动趋势及未来可能达到的水平。

1.1.2 预测的特点

1）预测与时间有关

预测涉及过去、现在和未来，过去、现在和未来都是时间概念，因此预测与时间直接关联。预测值只能是对预测对象在某一确定时间范围内的测定。预测间隔时间越长，预测结果越不准确。

2）预测依赖于信息

预测值寓于信息之中，这就要求预测工作者细致调查过去的历史资料并善于捕捉当前出现的苗头、征兆，进而对这些信息进行由此及彼、由表及里和去伪存真的加工。

3）预测的对象多为随机事件

预测研究的主要对象是随机事物。随机事物的发展永远存在不确定性，其可能发生，也可能不发生；可能这样发生，也可能那样发生。

4）预测的结果仅供参考

预测对象的随机性决定了预测结果的参考性。预测值与事物发生的未来结果一定会有误差。

1.1.3 预测的作用

科学预测，能使决策者了解未来，并使计划具有更高的预见性，把未来的不确定性问题，经定时、定性及定量分析，使未知降到最低限度，减少决策者的主观片面性，提高计划管理的科学性。通过图 1－1 可以看出预测在社会管理系统中的重要作用。

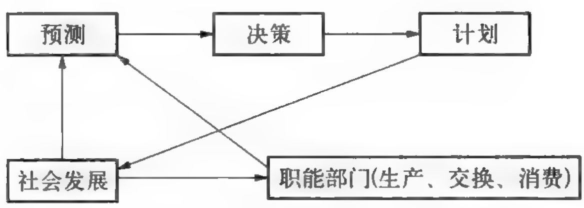


图 1－1 预测在社会管理系统中的作用

预测在当今社会的各个领域起着不可忽视的作用，具体表现在以下几个方面。

1）预测是编制经济和社会发展计划的前提

预测是认识未来的手段，它为研究政治、军事、外交、科技、经济、文化和教育提

供认识未来和发展的途径与措施。国家必须在科学预测的基础上制订国民经济发展规划。不经过科学预测就会错失良机，如美国在20世纪初到50年代，错误地选择了核发动机进行重点研究，结果苏联卫星首先上了天，美国因此错失了首次卫星上天的机会。

2）预测为正确确定计划目标和方向提供依据

预测能够揭示事物过去发展的客观规律和描述未来的发展变化趋势。在当今社会发展中，需要按客观规律、有计划地发展国民经济，并要在考虑未来发展的前提下去规划未来，制订出一整套切实可行的计划。科学预测能在时间、性质、数量上为计划提供具体的指标和说明，使计划更加切实可行，企业也必须在市场预测的基础上，制订生产和销售计划。

3）预测为选择最佳计划方案和最优决策提供依据

通过预测分析和使用不同的预测技术，可为决策者提供多种能相互替代的方案或途径，从而使上层机构或各级职能部门有条件地选择最佳计划方案和措施，作出正确的决策。如美国开发太空的计划、日本高速铁路技术的发展等都是以科学预测为依据的。

4）预测使企业经营管理科学化

预测，特别是经济预测和市场预测，在经营管理中的地位和作用是十分明显的。预测与企业经营管理的关系如图1-2所示。

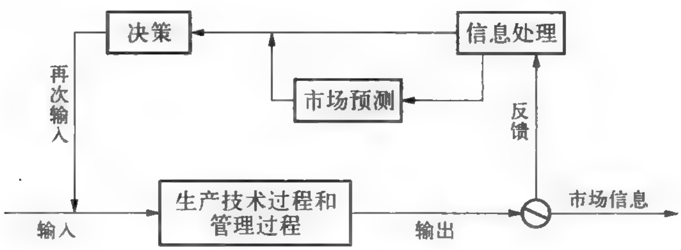


图1-2 预测与企业经营管理的关系

图1-2所示预测与企业经营管理的关系由两个子系统组成，一个是输入-输出系统，资源、劳力、指令、计划等输入后，经生产过程和管理过程向市场输出需求的产品。另一个是反馈系统，它是企业与外部环境联系，并且是评价其经营好坏的通道。输入-输出系统要由反馈系统通过市场上各种信息的迅速、准确搜集，并经过预测，来为决策者的正确决策提供重要情报和科学依据，进而确定再次输入的新内容（调整经营目标或作出相应的经营决策）。

预测在企业经营管理中发挥重要作用，使企业获得经济效益的例子不胜枚举。

据国外资料统计，利用预测得到的经济效益，比用于预测的开支高约50倍。

综上所述，预测的根本任务是通过科学方法的论证，为决策目标和决策方案的选择提供定性、定时及定量的科学依据，为决策部门制订社会经济发展规划、确定科研方向和重点提供科学依据。所以，预测工作任重而道远，甚至关系到人类生存和人类社会发展等重大问题。

1.2 预测的分类

根据研究任务的不同，按照不同标准可以有不同的分类。

1.2.1 按预测范围分类

按预测范围分类，可分为宏观预测与微观预测。

1）宏观预测

宏观预测，是指对整个社会、地区和部门的发展前景进行的各种预测。其特点是以整个社会、地区和部门的发展作为考察对象，研究相关指标之间的联系和发展变化，如国民经济总体规划预测、某一地区或行业的全面预测、某个港口未来发展前景预测等。

2）微观经济预测

微观经济预测是指对企业或具体单位的发展前景进行的各种预测。其特点是以企业或具体单位的发展前景为考察对象，研究各相关指标之间的关系和发展变化，如基层单位对产量、销量、利润、费用的预测，某个码头公司对吞吐量的预测等。

宏观预测与微观预测有密切关系，宏观预测应以微观预测为参考，微观预测应以宏观预测为指导，两者相辅相成。

1.2.2 按预测的时间长短分类

按预测的时间长短不同，可分为长期预测、中期预测、短期预测和近期预测。

1）长期预测

长期预测，是指对5年以上发展前景的预测。它是制订国民经济和企业生产经营发展的十年计划、远景计划，规定经济长期发展任务的依据。

2）中期预测

中期预测，是指对1年以上、5年以下发展前景的预测。它是制订国民经济和企业生产经营发展的5年计划，规定经济5年发展任务的依据。

3）短期预测

短期预测，是指对3个月以上、1年以下发展前景的预测。它是制订企业生产

经营发展年度计划、季度计划，明确规定短期发展具体任务的依据。

4）近期预测

近期预测，是指对3个月以下企业生产经营发展前景的预测。它是制订月、旬计划，明确规定近期活动具体任务的依据。

有人将短、近期预测合并，凡是1年以下的预测，统称为短期预测。

1.2.3 按预测方法的性质分类

按预测方法的性质不同，可分为定性预测和定量预测。

1）定性预测

定性预测，是指预测者通过调查研究，了解实际情况，凭自己的实践经验和理论、业务水平，对预测对象发展前景的性质、方向和程度作出判断、进行预测的方法，也称为判断预测或调研预测。预测目的主要在于判断预测对象未来发展的性质和方向，也可以在情况分析的基础上提出粗略的数量估计。定性预测的准确程度，主要取决于预测者的经验、理论、业务水平以及掌握的情况和分析判断能力。这种预测方法综合性强，需要的数据少，能考虑无法定量的因素。在数据不多或者没有数据时，可以采用定性预测方法，它和定量预测相结合，可以提高预测的可靠程度。

2）定量预测

定量预测，是根据准确、及时、系统、全面的调查统计资料和各种信息，运用统计方法和数学模型，对预测对象未来发展规模、水平、速度和比例关系的测定。由于定量预测和统计资料、统计方法有密切关系，也称统计预测。它包括时间序列预测和因果预测等。

定性预测比较简单易行，可利用有关人员的丰富经验、专业知识及掌握的实际情况，考虑不定量的因素的影响，进行比较切合实际的预测。其缺点在于，预测者由于工作岗位不同，掌握的情况不同，理论水平与实践经验各异，进行预测时包含的主观因素较多，往往会过分乐观而估计浮夸，或偏于保守而估计过低，对同一问题不同人会做出不同的判断，得出不同的结论。

定量预测，以调查统计资料和信息为依据，考虑各种现象发展的规律性和因果关系，建立数学模型，可以对预测对象未来发展前景进行科学的定量分析。其缺点在于，不考虑不定量的因素的影响，而是以基础条件和影响发展的各种主要因素比较稳定为前提，当这些条件和因素突然变动时，定量预测结果就会出现较大的偏误。

为了使预测结果比较切合实际，提高预测质量，为决策和计划提供可靠的依据，通常是将两种预测方法相结合，将定性预测结果和定量预测结果比较、核对，分

析其差异的原因，根据经验进行综合判断。利用定性分析对定量预测结果进行必要的修正和调整，才能取得良好的效果。

1.2.4 按预测的时态分类

按预测的时态分类不同，可分为静态预测和动态预测。

1）静态预测

静态预测，是指不包含时间变动因素，对同一时期预测对象因果关系的预测。

2）动态预测

动态预测，是指包含时间变动因素，根据预测对象发展的历史和现状，对其未来发展前景的预测。

1.3 预测与其他学科的关系

预测作为一门方法论的科学，它与经济决策、经济学、经济计划学、统计学、计量经济学、数理经济学、数学等多门学科有着密切的关系。预测是一门集各门学科于一身的交叉学科和边缘学科。

首先，预测和计划都是研究未来的，都是以经济学为理论基础的科学，但预测与计划的职能不尽相同。预测的主要职能是对未来的描述，计划的主要职能是对未来的部署。预测要说明的问题是如何发展下去、预测对象将来会变成怎样；计划要解决的问题是要使计划对象将来怎样做，现在应该采取何种措施和步骤。预测是计划的先导，计划是行动的安排。预测事先评审计划，预测对计划的作用在于：

（1）帮助我们认识和控制事物的不确定性，使对未来的无知降到最低限度。

（2）使计划的预期目标与可能变化的周围环境、经济变化保持一致。

（3）预测可事先了解计划实施后可能产生的结果以及事前应采取怎样的措施以帮助实现计划。

其次，预测是决策的依据，科学的预测是作出正确决策的保证。预测是决策的前提和手段。决策是从根本利益出发，遵循客观规律所作出的方针、政策或策略，是指导一切工作的依据。

就一个企业而言，预测、生产计划、决策可归纳为以下关系，如图 1－3 所示。

另外，预测与统计学、经济计量学、数理经济学、数学等也有密切的关系。预测以统计资料为依据，必然离不开统计学和统计工作提供准确、及时、系统、全面的统计资料。预测在研究预测对象的随机变化时离不开用概率论和数理统计的方法去揭示预测对象间的发展规律的数量关系。数理统计学的发展为预测提供科学的定

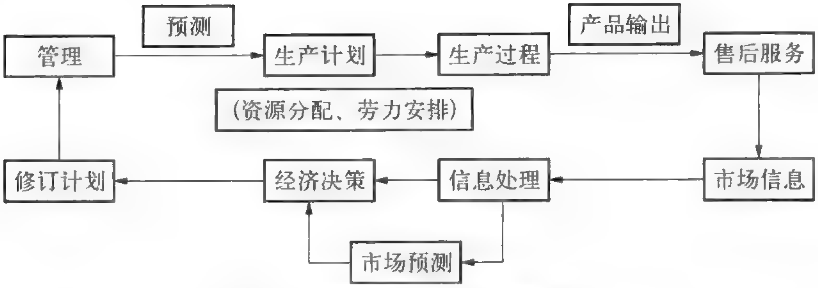


图 1-3　预测、生产计划、决策关系

量分析方法，从而也丰富和发展了预测的理论。

在预测中，定量预测是以数字为语言、数学模型为依据来研究和分析预测对象未来发展的数量关系。因此，在分析方法上、建模手段上必然离不开计量经济学、数理经济学与数学分析方法。一定的数学知识是保证定量预测数学化、模型化、科学化的前提。

数理经济学、经济计量学是经济学、数学、统计学的边缘学科，是用经济计量模型和投入产出模型来表达经济现象的运行规律；预测未来和规划未来的理论，定量预测的模型化必然离不开数理经济学与经济计量学的理论和方法，并以此作为专业方法论的基础。

总之，预测作为一门方法论学科，涉及社会科学、自然科学等各个领域。除以上提到的以外，还有近代飞速发展的信息论、系统论、控制论以及计算机工程等都丰富和发展了定量预测的手段、理论和方法，促进预测向数学化、模型化、科学化发展。

1.4　预测的步骤

为保证预测工作顺利进行，必须有组织有计划地安排其工作进程，以期取得应有的成效，为制定决策、编制计划和提高经营管理水平，提供有价值的情报。

预测程序一般包括以下基本步骤。

1.4.1　确定预测目的，制订预测计划

这是预测首先要解决的问题。确定预测目的，就是从决策与管理的需要出发，紧密联系实际需要，确定预测要解决的问题。预测计划是根据预测目的制订的预测方案，包括预测的内容、项目，预测所需要的资料，准备选用的预测方法，预测的进程和完成时间，编制预测的预算，调配力量，组织实施等。只有目的明确、计划科

学、有方向地安排预测内容、方法和工作进程,才能确定预测的经费和所需要的材料。预测无明确的目的、周密的计划,就会迷失方向,无所适从。

1.4.2 搜集、审核和整理资料

准确地调查统计资料和信息是预测的基础。进行预测需要掌握大量的数据和情况,掌握与预测目的、内容有关的各种历史资料,以及影响未来发展的现实资料,即要从多方面搜集资料。

资料按来源不同有内部资料和外部资料之分。内部资料,对公司和企业来说,是反映该单位历年情况的统计资料、市场调查资料和分析研究资料。外部资料,对公司和企业来说,是从本单位外部收集的统计资料和信息,政府统计部门公开发表和未公开发表的统计资料,兄弟单位之间定期交换的经济活动资料,报纸杂志上发表的资料,科学研究人员的调查研究报告,以及国外有关的信息和市场商情资料等。从这些资料中筛选出与本预测项目有密切关系的资料。筛选收集的标准有:直接有关性、可靠性、最新性。在把符合这三点的资料收集之后,经过研究,有必要时再搜集其他有关资料。

准确的资料,是提高预测准确性的前提条件之一。为了保证资料的准确性,要对资料进行必要的审核和整理。资料的审核,主要是审核来源是否可靠、准确和齐备,资料是否可比。资料的可比性包括:资料在时间间隔、内容范围、计算方法、计量单位和计算价格上是否保持前后一致。如有不同,应进行调整,使其前后一致才是可比的。资料的整理主要包括:对不准确的资料进行查证核实或删除;对不可比的资料调整为可比;对短缺的资料进行估计推算;对总体的资料进行必要的分类组合。

为了系统完备地为预测提供资料,应建立资料档案,系统地积累资料,以便连续地开展研究。只有根据预测目的和计划,从多方面收集必要的资料,经过审核、整理和分析,了解预测对象发展的历史和现状,认识其发展变化的规律性,才能使预测准确可靠,提高预测的质量。

1.4.3 选择预测方法和建立预测模型

在占有资料的基础上,进一步选择适当的预测方法和建立预测模型,这是预测准确与否的关键步骤。

对定性预测方法或定量预测方法的选择,应根据掌握资料的情况而定。当掌握资料不够完备、准确程度较低时,可采用定性预测方法。如对新的投资项目、新码头的发展进行预测时,由于缺少历史统计资料和信息,一般应采用定性预测方法,凭掌握的情况和预测者的经验进行判断预测。当掌握的资料比较齐全、准确程

度较高时，可采用定量预测方法。如对老码头生产经营的发展，可采用定量预测方法，运用一定的数学模型进行定量分析研究。为了考虑不定量因素的影响，在定量预测基础上要进行定性分析，经过调整才能定案。

在进行定量预测时，对时间序列预测法或因果预测法的选择，除根据掌握资料的情况而定外，还要根据分析要求而定。当只掌握与预测对象有关的某种统计指标的时间序列资料，并只要求进行简单的动态分析时，可采用时间序列预测法。当掌握与预测对象有关的、多种相互联系的统计指标资料，并要求进行较复杂的依存关系分析时，可采用因果预测法。

时间序列预测和因果预测都离不开数学模型，数学模型也称为预测模型，是指反映经济现象过去和未来之间、原因和结果之间相互联系和发展变化规律性的数学方程式。数学模型可能是单一方程，也可能是联立方程；可能是由一次方程组成的线性模型，也可能是由二次以上方程组成的非线性模型。预测模型选择是否适当，是涉及预测准确程度的一个关键问题。要建立数学模型，还必须估计模型参数（常数）。估计参数的方法，除传统的最小平方法外，还有多种专门的方法。不同的方法得出不同的参数估计值，影响预测出现不同的结果，应从实际出发，认真对待。

1.4.4 进行预测

在选择预测方法之后，即可进行定性预测或定量预测。定性预测时，由预测者对预测对象未来发展的性质、方向和程度作出判断。定量预测时，由预测者利用数学模型，由自变量估计因变量，给定一个自变量数值，估算出相对应的一个因变量数值，称为点估计；在既定概率保证下，估计实际值可能落在估计值上下的范围，称为区间估计。

1.4.5 计算、分析预测误差，评价预测成果

由于预测对象的发展变化受多种因素影响，预测是对预测对象发展前景的展望，难以预见到突发因素的影响，因而存在着不确定性，必有误差存在。预测误差是观察值与对应的预测值的离差。预测误差的大小反映预测的准确程度，预测误差越大，预测的准确程度越小，误差过大，就失去预测的应有作用。因此，必须计算预测误差，并分析产生误差的原因，把误差控制在一定范围内。同时还要和定性分析相结合，调整预测值，使预测结果尽量和实际情况接近。

评价定性预测成果时，主要是结合实际情况进行理论分析，评价预测是否切实可行。评价定量预测成果时，则要将各种假设检验与理论分析相结合，并与利用数学模型进行预测的成果相比较，以判断预测结果可信程度，看其是否接近实际。

1.4.6 改进预测方法，修正数学模型，提高预测质量

在预测过程中，随着时间的推移，要经常将已经出现的观察值与预测值相比较，以便发现情况的变化。当情况发生重大变化，原来的预测方法和数学模型不能如实地反映预测对象发展变化的实际时，就要修改预测方法和数学模型，以利于预测质量的提高。

1.5 预测的误差与精度分析

预测的误差就是指预测值与实际值之间的偏差程度。预测是一种根据历史和现在的资料对未来的估计和推算。由于所用资料的可靠性和完整性的问题，选用的预测方法适用程度问题，判断错误和计算误差以及情况突然变化等问题的影响，都会使预测数值与实际数值存在一定的差距，这个差距就是预测误差。预测的一个重要作用，就是可以计算预测结果的误差，并采用相应的数理统计方法进行分析，找出预测误差的变动规律，查明原因，采取措施，加以控制和改进，使预测结果更接近客观实际。

1.5.1 预测总会有误差

预测是根据过去和现在的数据研究找出其发展变化的规律。这些数据资料当然是已知的、确定的，但以所研究出来的规律去判断未来，必然会有误差。这是因为未来的问题具有未知性和不确定性。即使真实掌握了事物的演变规律，但由于预测对象未来会受到许多因素的干扰，预测结果往往会呈现趋向误差范围的上下摆动的趋势变化，即出现拐点。因此，预测除了研究预测对象的内在规律外，还应研究这种规律在处理预测时的误差范围和可能出现的拐点，以便使预测更好地反映未来和接近未来的真实情况。我们的任务就在于先计算出在一定概率保证下预测误差的范围，并改进预测方法或予以修正，以尽量减小误差。

1.5.2 预测误差产生的原因

1) 历史资料的可靠性

历史资料是预测的基础，资料的准确度如何，直接影响预测结果的正确性。在收集资料的过程中，会有以下因素产生误差：① 观测误差。如填报统计表有误、数字不准、内容不全、口径不一致等都属于观测误差。② 统计调查误差。如调查人员使用各种统计调查方法时，未能正确理解填报内容要求，或者报表的指标体系解释不清楚，抽样调查时未严格按抽样原则进行等。③ 资料整理加工误差。因为收

集资料并非实时收集处理，就会造成收集的资料在时间上不相对应，收集的资料不能反映某一时期的客观情况。

2）模型选择得当与否

客观事物的表现形态是各种各样的，表现的内在规律性也是不同的，故应针对不同的问题建立和使用不同的预测模型或方法。就是用同一统计资料进行预测，也因采用不同的方法，得出不同的结果。这样，模型或方法的选用得当与否必然影响预测误差。

3）预测对象成熟程度

预测对象发展的成熟程度低，今后发展的不确定性较大，预测误差也较大。

4）其他条件

预测人员的水平，进行预测所具备的条件，包括调查研究的条件、数据、资料、电子计算机等是否齐备，预测时间充足与否等，都对预测误差有一定影响。

思考与练习

1. 什么是预测？为什么要进行预测？
2. 预测有哪些特点？
3. 预测与决策、计划有何关系？
4. 预测的分类方法有哪些？
5. 简述预测的基本步骤。

2 定性预测法

2.1 定性预测概述

2.1.1 定性预测的概念和特点

定性预测是指预测者依据熟悉业务知识、具有丰富经验和综合分析能力的人员与专家，根据已掌握的历史资料和直观材料，运用个人的经验和分析判断能力，对事物的未来发展做出性质和程度上的判断；然后再通过一定的形式综合各方面的意见，作为预测未来的主要依据。

定性预测的特点在于：

(1) 着重对事物发展的性质进行预测，主要凭借人的经验以及分析判断能力。在对预测对象所掌握的历史统计资料不多，或影响因素复杂，难以分清主次，或对主要影响因素难以定量分析等情况下，定性分析方法将是适用性很强的方法。

(2) 着重对事物发展的趋势、方向和重大转折点进行预测。例如，在港口管理方面，它主要适用于下列情况的预测：港口货物吞吐量预测、临港产业发展趋势研究及预测、港口集疏运系统运输能力的可持续性、进出港船舶交通量预测等。

定性预测的方法很多，本书侧重于德尔菲法和主观概率法的介绍。

2.1.2 定性预测和定量预测

1) 定性预测的优缺点

定性预测的优点在于：注重事物发展在性质方面的预测，具有较大的灵活性，易于充分发挥人的主观能动性，且简单迅速，省时、省费用。

其缺点是：易受主观因素的影响，比较注重人的经验和主观判断能力，从而易受人的知识、经验和能力的多少、大小的束缚和限制，尤其是缺乏对事物发展做数量上的精确描述。

2) 定量预测的优缺点

定量预测的优点在于：注重事物发展在数量方面的分析，重视对事物发展变化的程度做数量上的描述，更多地依据历史统计资料，较少受主观因素的影响，可以利用电子计算机对相关统计方法和数学方法做大量的计算处理。

其缺点是：比较机械，不易灵活掌握，对信息资料的质量和数量要求较高，而且不易处理有较大波动的信息资料，更难以预测事物的变化。

2.1.3 定性预测和定量预测之间的关系

定性预测和定量预测并不是相互排斥的，而是可以相互补充的，在实际预测过程中，应该把两者正确地结合起来使用。

具体地说，就是在占有比较完备的统计资料的条件下，可以先用一定的统计方法进行加工处理，找出有关变量之间的规律性联系，作为预测未来的一个重要依据。但是，任何数学方法或统计方法的应用，都是以过去的信息资料为基础来预测未来，如果在预测期内发生了重大变化，出现了新的重大影响因素，如政府或管理部门的政策、方针有重大改变，区域经济发展趋势的重大改变，市场经营战略或市场经营组合有重大改变，或出现了过去的信息资料所没有反映的其他重要情况，则应根据以上新产生的因素，对定量预测方法所得到的结果加以修正。这就需要依靠熟悉情况和业务的人员和专家，运用定性预测方法，提出修正意见。而在使用定性预测方法的同时，也要尽可能地采用数学方法，对事物发展变化的趋势、方向、程度和转折点出现的时间等做出数量上的测算，以保证预测结果的客观准确。

可见，在实际的预测工作中，只有把定性预测方法和定量预测方法正确地结合起来，相互补充、相互检验和修正，才能取得较好的预测效果。

2.2 德尔菲法

2.2.1 德尔菲法的含义

德尔菲法(Delphi)是根据有专门知识的人的直接经验，对研究的问题进行判断、预测的一种方法，也称专家调查法。它是美国兰德公司于 1964 年首先用于预测领域的。该方法是采用函询调查方法，对所预测问题有关领域的专家分别提出问题，并将其意见综合、整理、汇总、反馈，经过多次反复循环，最后得到一个比较一致的且可靠的意见。

2.2.2 德尔菲法的特点

德尔菲法的特点可以归纳为以下几点：

(1) 匿名性。为克服专家易受心理因素影响的不足，德尔菲法采用匿名函询征求意见。应邀参加预测的专家之间不发生联系，消除了心理因素的影响。专家可以参考前一轮的预测结果修改自己的意见，而无需公开说明，无损自己的

威望。

（2）反馈性。德尔菲法一般要经过几轮的函询与反馈。预测小组对每一轮各位专家的结果作出统计和处理，作为反馈材料再寄给每位专家，达到相互启发的目的。

（3）统计性。为了给出定量预测的结果，德尔菲法采用统计方法处理每一轮的专家意见，使预测结果具有统计性特点。

2.2.3 德尔菲法的预测程序

此种方法的具体工作步骤如下：

1）拟定调查课题

由预测组织者拟定出需要预测的课题，列成调查表，并提供有关背景资料。一次调查的问题不宜过多、过杂。

2）组成专家调查组

所选择的专家应具有与预测课题有关的专业知识、工作经验、预测分析能力和一定的声望，是在预测问题所属专业领域内有丰富实践经验的实际工作者或有较深造诣的理论研究者。而且他们对预测的问题有热心、有兴趣，愿意参加并能胜任。挑选的专家们应在专业水平、年龄、职务、性格、社会背景等方面具有广泛的代表性，以便取得较全面的信息。专家的人数由十几人到几十人不等，具体人数应视预测课题的复杂性而定。

3）进行第一轮调查

主持预测单位一方面将调查表发给每位专家，另一方面也可以把与征询问题有关的各种材料寄给每一位专家；同时，请专家提出自己还需要什么资料，尽量满足专家的要求。专家们则根据通知的要求，对所预测事物提出个人的判断与分析，并说明依据与理由，在一定的时间内寄回自己的初步意见。

调查表是专家们回答问题的主要依据，所以，拟定调查表，要用前言说明预测的目的和任务，回答问题者应注意的事项，寄回调查表的时间规定等，并示范说明如何回答问题。调查表中的问题要集中，要有针对性，要使各个项目构成一个有机整体。同时，调查表提出的问题也不要过多，一般认为在25个左右为宜，提出的问题应尽量避免使用“普遍”“广泛”“正常”等缺乏定量概念的用语。调查表的回答要采用简练的方式，如填上数字、日期、同意（√）、不同意（×）等，有利于专家集中精力思考问题，也可在调查表中留出适当余地，便于专家阐明自己的看法和意见。

4）进行逐轮调查

主持预测单位把收集的专家意见加以集中整理，将不同的预测结果及其依据与理由，以匿名的方式再分送给各专家，进行第二轮征询，要求专家补充、修改各自

的预测，并加以说明或评论。同样，将第二轮征得的专家的意见进行汇总整理，再反馈给每位专家，进行第三轮征询，如果意见比较集中，进行到第三轮调查即可结束；否则，还要进行第四轮、第五轮的调查。

5）得出预测结论

专家根据各方面的资料、数据、意见，提出自己的补充或修改预测意见，并说明其依据与理由。这种不记名的反复征询，一般经过四五轮，意见便逐渐趋向一致，最后得出比较切合实际的集中答案。

运用德尔菲法进行预测的程序如图 2－1 所示。

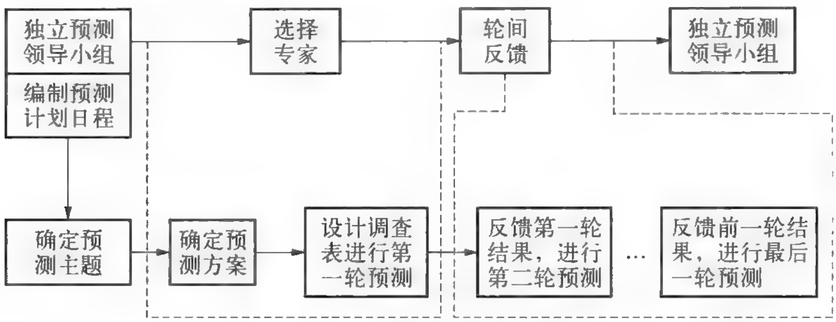


图 2－1 德尔菲法预测程序

这里需要指出，各轮征询所得到的专家意见，要利用统计方法进行处理。依据问题的不同类型，采用不同的统计处理法加以汇总归类。

（1）当专家们的意见是一系列可以比较大小的数据或有时间先后顺序时，往往用中位数代表专家们预测的协调结果，用上、下四分点代表预测的分散程度。

如果对专家们的预测结果按照由大到小的顺序排列（时间按先后，数量按大小）。X 表示专家所提供的预测数据，当有 N 个专家时，共有 n 个相异的预测数据，排列顺序为

$$X_1 < X_2 < \cdots < X_{n-1} < X_n$$

其中中位数的计算公式如下：

$$M_d = X_{k+1} \qquad n = 2k+1\ (\text{奇数})$$

$$M_d = \frac{X_k + X_{k+1}}{2} \qquad n = 2k\ (\text{偶数})$$

式中，M_d 为中位数；X_k 为第 k 个数据；X_{k+1} 为第 $k+1$ 个数据；k 为正整数。

当 $n = 2k+1$（奇数）时，除去第 $k+1$ 项后，该序列分作两部分，采用上述求中

位的公式，求出前 k 项的中位数，为原序列的下四分点；求出后 k 项的中位数，为上四分点。当 $n=2k$（偶数）时，按照同样的方法求出前 k 项的中位数，为原序列的下四分点，求出后 k 项的中位数，为原序列的上四分点数。

【例 2-1】 某集装箱码头购买新型船舶装卸桥 4 台，聘请 16 位专家对购买新机械后该码头的年吞吐量进行预测，16 位专家在最后一轮的预测值如表 2-1 所示。

表 2-1 某集装箱码头年吞吐量 单位：万 TEU

专家编号	预测吞吐量	中位数和上、下四分位数
1	12.5	
2	13.8	
3	14.0	
4	14.0	下四分位数 14.0
5	14.0	
6	14.5	
7	14.7	
8	15.0	中位数 15.0
9	15.0	
10	15.0	
11	15.1	上四分位数 15.4
12	15.3	
13	15.5	
14	15.8	
15	16.2	
16	16.5	

这里，$n=16$ 是偶数，则 $k=n/2=8$ 时，中位数 M_d 是第 8 位数与第 9 位数的平均数。

$$M_d=(X_8+X_9)/2=(15.0+15.0)/2=15.0\ (\text{万 TEU})$$

上四分点的数位 $k+k/2$ 与 $k+k/2+1$ 项数值的平均数，即第 12 项、13 项数值的平均数。

$$Q_{\text{上}}=(X_{k+k/2}+X_{k+k/2+1})/2=(15.3+15.5)/2=15.4\ (\text{万 TEU})$$

下四分点的数为 $k/2$ 与 $k/2+1$ 项的数值的平均数，即第 4 项与第 5 项数值的平均数。

$$Q_{下} = (X_{k+k/2} + X_{k+k/2+1})/2 = (14.0 + 14.0)/2 = 14\ (万\ \mathrm{TEU})$$

式中，M_d 表示了预测数值的分布中心，代表专家的预测意见，也就是专家们的期望值；$Q_{下}$ 表示预测区间的下限；$Q_{上}$ 表示预测区间的上限，上下限之间的范围表示预测区间。由于专家意见趋于中位数，有 50%的专家预测值在预测区间以内，该案例表示该新产品的预测销售值为 15 万 TEU，有 50%的专家预测者在(14～ 15.4)万 TEU。

(2) 利用平均数，求出预测值。

【例 2-2】 某物流公司要开展新的物流运输服务，为了在正式开展之前，摸清该种运输服务的需求情况，他们采用德尔菲法进行预测，其做法是首先选取了包括业务人员、推销人员在内的 9 位非常熟悉市场和业务情况的人员；其次将该种运输服务特点、用途以及同类运输服务的价格和需求量单独向他们做了介绍，并发给他们意见书，邀请他们分别做出个人的分析判断；然后，将上述判断汇总整理，再匿名返回给每位人员，如此反馈 3 次，结果如表 2-2 所示。

表 2-2 第三方物流公司运输服务需求预测 单位：千吨/年

专家小组成员	第一次判断需求量			第二次判断需求量			第三次判断需求量		
	最低需求量	最可能需求量	最高需求量	最低需求量	最可能需求量	最高需求量	最低需求量	最可能需求量	最高需求量
A	25	60	70	25	65	75	25	70	77
B	30	55	70	30	50	80	31	55	80
C	40	60	75	35	55	70	34	64	70
D	40	55	85	30	60	80	40	60	78
E	20	40	80	30	55	70	30	56	70
F	20	40	45	25	50	60	26	52	58
G	30	50	55	35	40	55	34	43	56
H	18	25	30	20	30	40	22	32	40
I	20	36	50	20	30	40	20	34	42
平均							29.1	51.8	63.4

如按简单平均数计算，根据第三次预测，该物流服务需求量为

$$X = (29.1 + 51.8 + 63.4)/3 = 48.1\ (千吨/年)$$

若将最低需求量、最可能需求量、最高需求分别按 0.2，0.5 和 0.3 的权数进行

加权平均，预测该物流服务需求量为

$$X = 29.1 \times 0.2 + 51.8 \times 0.5 + 63.4 \times 0.3 = 50.7\text{（千吨/年）}$$

德尔菲法的主要优点：由于专家调查法具有匿名的性质，可以保证信息的交流不受权威、资历、口才等原因的影响；由于调查是循环反复进行多次，可以使每个人不仅知道集体答案的分布情况，并能了解持有不同意见者的理由，促使被调查者可充分进行思考和修正自己的意见；由于预测结果是综合全体专家的意见，因而可使预测结果具有较大的可靠性和权威性。这种方法应用面比较广，在资料不甚全面或不够多的情况下均可采用。

2.2.4 德尔菲法的优缺点

德尔菲法的优点在于：

(1) 可以加快预测速度，节约预测费用。

(2) 可以获得各种不同但有价值的观点和意见。

(3) 在历史资料不足或不可测因素较多时较适用。

德尔菲法的缺点在于：

(1) 对于分地区的顾客群或产品的预测可能不可靠。

(2) 责任比较分散。

(3) 专家的意见有时可能不完整或不切合实际。

2.3 管理、销售人员意见调查预测法

此法又名集合意见法或经验判断预测法。其特点是以领导层和基层业务人员的经验和判断为基础，经过分析综合，以判断未来的市场情况。此法的优点是：在短时间内能集中有关人员的意见，迅速做出判断，简单易行。这种方法在缺乏预测资料时特别有用，如果预策者有较丰富的经验和较强的分析判断能力，并且对各方面的情况比较熟悉，就可以得到较好的预测结果。此法的缺点是主观意志较多，客观的数据和资料不足，容易发生偏差。例如，基层销售人员的意见，多从完成销售任务出发，对市场估计偏低。同时，判断预测多用于一般性预测，内容不够细致，对市场变化、需求意向等难以细分。判断预测法可细分为下列几种形式。

2.3.1 经营管理人员意见调查预测法

经营管理人员意见调查预测法，就是由企业的负责人把与市场有关或者熟悉

市场情况的各种负责人员和中层管理部门的负责人召集起来，让他们对未来的市场发展形势或某一种大市场问题发表意见，做出判断；然后，将各种意见汇总起来，进行分析研究和综合处理，对市场需求的情况做出主观判断，最后得出市场预测结果。参加判断的人员，主要是企业中主管经营业务的经理和有关部门主管干部。主管人员在预测时，应充分利用企业现有资料，熟悉市场情况，掌握信息动态，对于不同的预测结果，应在相互讨论、相互比较的基础上，修正原始意见，做出一个较可靠的估计。

这种方法的预测步骤如下：

1）企业经营、管理人员分别提出自己的预测方案

各调查预测人员在提出预测方案的过程中，要进行三个方面的分析和判断。

首先，要进行以下几方面质的分析：① 历史趋势，目前市场的动态，经济、贸易、航运市场等发展的新变化，货源的流动情况；② 各企业(单位)的产品(服务)适销对路的情况，商品(服务)资源、流通渠道以及投入市场的可能性；③ 流动资金的来源和运用情况、生产设备库存结构以及生产目标；④ 劳动的组织情况，经营管理的改善措施及可能达到的结果。

其次，在质的分析的基础上，进行三方面量的分析：① 可能出现的自然状态，即根据分析，判断市场未来会出现的几种销售可能性；② 由于各种自然状态出现的可能性大小不同，需要分别确定其出现的概率；③ 确定各种自然状态下的估计值。

最后，计算在各种自然状态及概率下期望达到的预测值，在期望的基础上确定出各自的预测方案。

【例 2-3】 假定某集装箱码头为了预测明年船舶装卸量的市场需求情况，邀请三名主管，分别是商务部经理、生产部经理和财务部经理，让他们分别做出明年码头吞吐量的预测方案，他们的预测情况如表 2-3 所示。

表 2-3 经理人市场需求预测表 单位：万吨

经理	市场需求状态	估计值	概率	期望值
商务部经理	最高市场需求	300	0.2	260
	最可能市场需求	260	0.6	
	最低市场需求	220	0.2	
生产经理	最高市场需求	360	0.3	328
	最可能市场需求	320	0.5	
	最低市场需求	300	0.2	

续 表

经 理	市场需求状态	估计值	概 率	期望值
财务经理	最高市场需求	320	0.3	300
	最可能市场需求	300	0.4	
	最低市场需求	280	0.3	

对于三位经理的预测期望值，由表 2－3 中一一列出，对表中三位经理的三个预测方案，也可以综合为一个预测方案计算出其综合预测值，如按相同的权数，则他们的综合预测值为

$$(260+328+300)\div 3=296(\text{万吨})$$

2）对三位经理的预测方案加以综合，利用权重分析判断，确定企业的预测方案

在确定其预测值时，一般根据他们对市场环境变化等情况的了解程度不同，给予不同的权重。一般来说，企业商务部经理对企业产品的市场需求情况最为了解，所以给的权重相应较大；而企业生产经理长期在生产第一线，对市场需求情况了解较少，所以就给予较少的权重；至于财务经理，由于长期从事财务管理，其预测值也比较能客观地反映市场需求实际。所以，我们假设商务部经理的预测取权重 5，生产经理取权重 2，财务经理取权重 3，采用加权平均法计算下一年度的吞吐量预测值，则有

$$(260\times 5+328\times 2+300\times 3)\div(5+3+2)=285.6(\text{万吨})$$

3）对企业的预测方案做出必要的调整

由上述预测方案可以看出，经过加权平均的吞吐量综合预测值，虽然高于商务部经理的预测方案（260 万吨），但却低于生产经理的预测方案（328 万吨）、财务经理的预测方案（300 万吨）。这是因为：商务部经理是吞吐量预测方案的直接执行者，码头的市场需求方案与他们的经济利益有直接的关系，因此，一般他们的预测值较低。由此形成的预测方案并不是码头的最后方案，码头还应当通过对影响市场需求的各种因素和码头的主观条件进行综合分析、推理、判断，并召集一定的会议集思广益，对综合的预测值进行适当调整，才能作为码头最后确定的市场需求预测方案。

这种方法的优点在于：

（1）迅速、及时和经济，不需要经过复杂的计算，也不需要多少预测费用，就可以及时得到预测结果。

（2）集中了解各个方面熟悉市场情况的有经验中高级管理人员的意见，可以

发挥集体的智慧，使预测结果比较可靠。

(3) 使用这种方法不需要大量的统计资料，更适合于那些不可控因素较多的预测对象。

(4) 如果市场发生了变化可以自己进行修正。

这种方法的缺点在于：

(1) 预测结果容易受主观因素影响。

(3) 对市场变化、客户的愿望等问题了解不细，因此预测结果一般化。

2.3.2 销售人员意见调查预测法

销售人员意见调查预测法就是组织者召集有经验的销售人员对客户的需求量、市场需求变化趋势、竞争对手动向等问题进行预测，然后对预测结果进行综合的预测方法。由于销售人员一般都很熟悉市场情况，因此，这一方法具有一些显著的优势。

此预测方法参加的人员是企业基层与商务有关的业务人员。基层业务人员在第一线工作，直接接触商品(服务)的消费者，预测判断结果较具体，贴近实际，但由于接触面窄，容易出现以偏概全的情况，在实际预测过程中应注意纠正。

【例 2-4】 一个煤炭码头有 3 名销售人员(甲、乙、丙)，对明年的煤炭装卸量的市场需求情况预测如表 2-4 所示。

表 2-4 煤炭码头销售人员预测表 单位：万吨

销售人员	装卸量状态	估计值	概 率	期望值
甲	最高装卸量	210	0.2	161
	中等装卸量	160	0.5	
	最低装卸量	130	0.3	
乙	最高装卸量	200	0.3	175
	中等装卸量	170	0.4	
	最低装卸量	150	0.3	
丙	最高装卸量	180	0.3	156
	中等装卸量	150	0.6	
	最低装卸量	120	0.1	

表 2-4 中列出了每名销售人员的预测期望值，如果认为每个销售人员的预测比较符合实际，那么采用算术平均法可以计算出该码头下一年的装卸量预测值：

$$装卸量预测值 = (161 + 175 + 156) \div 3 = 164(万吨)$$

也可以采用加权平均法计算预测值，对不同的销售人员给予不同的权重。如果需要修正，也可以对某些销售人员的预测值进行修正。在实际的调查预测中，可以调查更多的销售人员，也可以选取本企业以外的销售人员进行销售预测。

2.3.3 综合判断法

这是综合主管人员、基层业务人员及其他有关方面的判断而确定的预测结果，它首先由企业负责人召集销售、计划、生产、财务等部门的负责人或销售人员广泛交换意见，预测生产量，然后，将不同人员的预测值进行综合，得出预测结果。因为各类人员所处的工作环境不同，判断各有优缺点，如能全面综合，则预测效果会更好。

【例 2-5】 假定某码头公司 3 名高层管理人员和 3 名销售人员分别对下年的生产量进行预测，并计算出他们的预测期望值，然后根据他们的预测期望值进行加权平均，得出该码头下一年的生产预测值。

将经营管理层（商务部经理、生产经理、财务经理）和销售人员的期望值进行综合分析判断。因这几类人员的预测意见的重要性不同，故在综合时，采用加权平均法。如果商务部经理的权重为 4，生产经理的权重为 2，财务经理的权重为 3，销售人员的权重为 1，对他们的生产量预测进行加权平均，则得

$$(260 \times 4 + 328 \times 2 + 300 \times 3 + 164 \times 1) \div (4 + 2 + 3 + 1) = 275.4(万吨)$$

2.4 主观概率预测法

2.4.1 主观概率法的概念

主观概率法是对德尔菲法、管理销售人员意见调查预测法的不同定量估计，进行集中整理的常用方法。主观概率是指在一定条件下，个人对某一事件在未来发生或不发生可能性的估计，反映个人对未来事件的主观判断和信任程度。

主观概率也必须符合概率论的基本公理，即每一事件发生的概率大于等于零，小于等于 1；必然发生的事件概率等于 1；必然不发生的事件概率等于零；两个互斥事件之和的概率等于它们的概率之和。

客观概率，是指某一随机事件经过反复试验后，出现的频数，也就是对某一随机事件发生的可能性大小的客观估量。如掷一枚硬币，出现国徽面和出现数字面的客观概率各为 1/2。

客观概率与主观概率的根本区别在于，客观概率具有可检验性，主观概率则不具有这种可检验性。

在有些现象无法通过试验确定其客观概率，或由于资料不完备无法计算客观概率时，常常采用主观概率法进行预测。常用的主观概率法有：主观概率加权平均法和累计概率中位数法。

2.4.2 主观概率加权平均法

这种方法是以主观概率为权重，对各种预测意见进行加权平均，求得综合性预测结果的方法。其步骤如下：

第一步：确定主观概率。根据过去预测的准确程度来确定各种可能情况的主观概率。

第二步：计算综合预测值。以某物流公司的统计人员和计划人员对下一年首季收入额的预测为例：

首先，以主观概率为权重，计算每人预测的最高收入、最低收入和最可能收入的加权算术平均数，作为个人预测期望值。各统计数据如表 2－5 所示。

表 2－5 统计人员预测期望值统计表

统计员	估 计	收入额/万元	主观概率	收入×概率
(1)	(2)	(3)	(4)	(5)
甲	最高收入	1 000	0.3	300
	最可能收入	800	0.5	400
	最低收入	600	0.2	120
	期望值			820
乙	最高收入	1 200	0.2	240
	最可能收入	1 000	0.6	600
	最低收入	800	0.2	160
	期望值			1 000
丙	最高收入	900	0.2	180
	最可能收入	700	0.5	350
	最低收入	500	0.3	150
	期望值			680

如统计员甲的期望值为

$$(1\,000 \times 30 + 800 \times 50 + 600 \times 20) \div 100 = 820(\text{万元})$$

其次，以主观概率为权数，计算各个期望值的平均数。

如果三位统计员的判断能力不相上下，其主观概率各为 1/3，则三人预测的平均收入额为

$$(820 + 1\,000 + 680) \div 3 = 833.33\ (\text{万元})$$

如果计划员甲的期望值为 950 万元，乙的期望值为 750 万元，两人的主观概率各为 50%，则计划员预测的平均销售额为

$$(950 + 750) \div 2 = 850\ (\text{万元})$$

如果统计员的主观概率为 60%，计划员的主观概率为 40%，则该公司明年首季的预测收入额为

$$(833.33 \times 60 + 850 \times 40) \div 100 = 840\ (\text{万元})$$

上述每人期望值的主观概率，主要根据过去每人判断预测的准确程度确定。

第三步：计算平均偏差程度，校正预测结果。将过去若干季的实际数和预测数对比，计算比率、平均比率和平均偏差程度。设过去 8 个季度的实际数与预测数之比如表 2－6 所示。

表 2－6　实际数与预测数之比

季　数	1	2	3	4	5	6	7	8	平均比率
实际数/预测数	0.98	1.03	1.02	0.86	0.97	1.01	0.93	1.04	0.98

平均比率是各季比率的简单算术平均数，为 98%，即实际数比预测数有高有低，平均为 98%。则平均偏差程度为 98%－1＝－2%，这说明实际数比预测数平均低 2%，也就是预测数比实际数平均偏高 2%。因此应将预测数扣除 2%进行校正，则经校正后的该公司下一年首季预测收入额为 840×98%＝823.2(万元)。

2.4.3　累计概率中位数法

这种方法是根据累计概率确定不同预测意见的中位数，对预测值进行点估计的区间估计方法，其步骤如下：

第一步：确定主观概率及其累计概率。

现以某物流公司的流通费用率为例。某公司根据 2010 年 1 月至 2012 年 6 月共 30 个月的流通费用率，预测 2012 年 12 月的流通费用率。其统计资料如表 2－7 所示。

表 2-7 2010 年 1 月至 2012 年 6 月各月流通费用率

年＼月	1	2	3	4	5	6	7	8	9	10	11	12
2010	3.3	2.8	2.9	3.0	2.9	3.0	3.4	3.1	3.0	2.9	3.0	3.1
2011	3.5	3.4	3.3	3.4	3.5	3.5	4.6	4.0	4.0	4.4	4.9	6.3
2012	8.3	8.3	9.6	8.8	8.3	9.4						

从表 2-7 资料可以看出，2011 年 12 月以来流通费率有明显的迅速上升趋势。要外推预测 2012 年 12 月份的流通费率，可用意见征询表进行调查，确定主观概率及其累计概率和有关预测资料。意见征询表的目的是，要得到能用来预测 2012 年 12 月份流通费率的信息。每个预测者要根据自己对流通费率未来发展趋势的认识对其主观概率做出估价。

可通过征询表向参与预测者提出一系列问题，预测者回答时要符合概率基本合理的要求，并要参照如图 2-2 所示标尺回答问题。

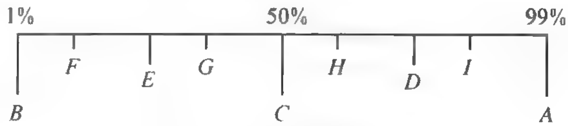


图 2-2 参照标尺

标尺上除两端的两点外，中间各点都代表随机变量样本空间的 12.5%。上列标尺代表从 1%至 99%的累计概率区间。利用标尺可以了解每一答案在累计概率区间的位置。

通常征询表中所提的问题如下：

A. 你认为 2012 年 12 月流通费率的最高值可能是多少？所谓最高值，就是实际流通费率有 99%的概率小于或等于这个值。

B. 你认为 2012 年 12 月流通费率的最低值可能是多少？所谓最低值，就是实际流通费率小于或等于这个值的概率只有 1%。

C. 你在 A、B 答案之间确定一个流通费率，要求在其以上或以下出现的实际值的概率各为 50%，这是概率分布的中位数。

D. 你在中位数 C 和最高值 A 之间，确定一个能把这一区间分成等概率的两半的流通费率。这是全数列的上四分位数的数值。实际值小于或等于这个值的概率为 75%。

E. 在最低值 B 和中位数 C 之间，确定一个等概率值。这是全数列的下四分

位数的数值。实际值小于或等于这个值的概率为25%。

F. 在 B 与 E 间确定一个等概率值，小于或等于这个值的概率为12.5%。

其余各点的概率类推。

将上述意见征询表发给每个预测组成员填好答案收回后，即可得到各种意见及其主观概率和累计概率。

第二步：汇总整理意见征询表，进行点估计和区间估计。

设预测者有12人，各人对每一问题的答案都汇总列入表2-8中。为了形成整个预测组的统一的累计概率分布函数，要把从 A 到 B 的9个点中每一个点的12个估计值加以平均。将所得累计概率的分布函数的中位数，确定为2012年12月份流通费率的点估计值，如表2-8所示。

表2-8 意见征询表的答案汇总表

预测者编号	累积分布函数沿横轴的点								
	B	F	E	G	C	H	D	I	A
1	6.0	6.25	6.50	6.75	7.0	7.25	7.50	7.75	8.0
2	6.0	6.40	6.50	7.00	8.3	8.40	8.50	9.40	9.5
3	8.0	8.13	8.25	8.38	8.5	8.63	8.75	8.88	9.0
4	6.0	6.70	7.50	8.00	8.0	8.60	8.70	8.80	9.0
5	5.0	5.50	6.00	6.50	7.5	8.00	8.25	8.50	9.0
6	8.0	8.23	8.45	8.68	8.9	9.13	9.35	9.58	9.8
7	7.8	8.00	8.20	8.50	8.8	9.00	9.30	9.40	9.6
8	8.0	8.20	8.40	8.60	8.8	9.00	9.20	9.40	9.6
9	7.2	7.80	8.26	8.40	8.6	8.80	9.20	9.60	10.0
10	6.0	6.68	8.25	8.38	8.5	8.63	8.75	9.66	10.0
11	9.2	9.25	9.30	9.35	9.4	9.45	9.50	9.70	9.8
12	6.5	6.80	7.20	8.10	8.8	9.00	9.10	9.30	9.5
平均数	6.98	7.33	7.73	8.05	8.43	8.49	8.84	9.14	9.40
累计概率/%	1.0	12.5	25.0	37.5	50.0	62.5	75.0	87.5	99.0

从表2-8中得知2012年12月份流通费率的中位数为8.43之外，同时还得

到其他信息，如根据预测者的估计平均数，2012 年 12 月份流通费率低于或等于 8.84的概率为 75%。

根据表 2 - 8 各点的平均数和累计概率，可以绘制流通费率的累计概率分布函数图，如图 2 - 3 所示。

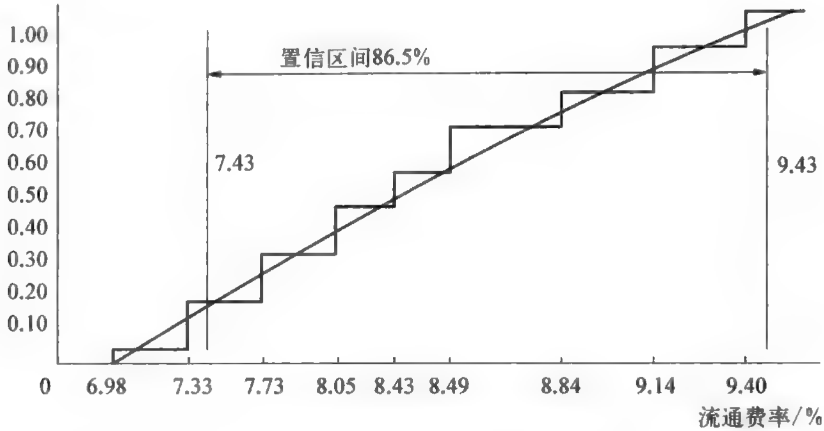


图 2 - 3 流通费率累计概率分布函数图

为了外推与插补用，可将图 2 - 3 折线修匀成一条平滑曲线。设本例要求预测误差不超过 1%，即以点预测为中心，置信区间为±1%，也就是实际值在点预测的±1%范围内，则 8.43%±1%为[7.43%，9.43%]。根据图 2 - 3 可以算出这个区间的概率为 0.99—0.125=86.5%。以这个概率保证流通费率预测的置信区间为[7.43%，9.43%]，可靠程度是相当高的。

第三步：计算预测误差，校正预测值。

利用主观概率法进行判断预测，预测的准确程度和预测误差的大小成反比。为了提高预测的准确程度，对点估计值加以校正，应计算预测误差。设某公司用上述预测方法进行预测已有一年，可比较预测值与观察值的误差，计算以下几个平均误差指标，对预测值进行校正。计算数据参看表 2 - 9。

表 2 - 9 计算数据表

月 份	观察值 x	预测值 $\hat{x}$	$x-\hat{x}$	$\lvert x-\hat{x}\rvert$	校正值 $\hat{x}'$	$\lvert x-\hat{x}'\rvert$
(1)	(2)	(3)	(4)=(2)—(1)	(5)	(6)=(3)+$\bar{e}$	(7)=(2)—(6)
2011 年 7 月	4.6	5.4	—0.8	0.8	5.867	1.267
8 月	4.0	4.2	—0.2	0.2	4.667	0.667

续 表

月　份	观察值 x	预测值 $\hat{x}$	$x-\hat{x}$	$\|x-\hat{x}\|$	校正值 $\hat{x}'$	$\|x-\hat{x}'\|$
(1)	(2)	(3)	(4)=(2)−(1)	(5)	(6)=(3)+$\bar{e}$	(7)=(2)−(6)
9月	4.0	3.9	0.1	0.1	4.367	0.367
10月	4.4	4.0	0.4	0.4	4.667	0.067
11月	4.9	4.8	0.1	0.1	5.267	0.367
12月	6.3	6.5	−0.2	0.2	6.967	0.667
2012年1月	8.3	8.9	−0.6	0.6	9.367	1.067
2月	8.3	6.1	2.2	2.2	6.567	1.733
3月	9.6	7.3	2.3	2.3	7.767	1.833
4月	8.8	7.7	1.1	1.1	8.167	0.633
5月	8.3	7.5	0.8	0.8	7.697	0.333
6月	9.4	9.0	0.4	0.4	9.467	0.067
合计	80.9	75.3	5.6	9.2	80.9	9.068

(1) 预测平均误差

$$\bar{e}=\frac{\sum(x-\hat{x})}{n}=\frac{5.6\%}{12}=0.467\%$$

$\bar{e}$ 是测定预测的偏向性的统计指标。上例表明预测有系统误差，偏低0.467%，可用 $\bar{e}$ 来校正原预测值 $\hat{x}$，得校正预测值 $\hat{x}'$，填入表2-9第(6)栏。

$$\hat{x}'=\hat{x}+\bar{e}=\hat{x}+0.467\%$$

如2011年7月的校正预测值为

$$5.4\%+0.467\%=5.867\%$$

(2) 平均绝对误差

$$|\bar{e}|=\frac{\sum|(x-\hat{x})|}{n}=\frac{9.2\%}{12}=0.77\%$$

$|\bar{e}|$是测定预测准确性的统计指标。

(3) 校正预测的平均绝对误差$|\bar{e}'|$是测定校正后预测一致性的统计指标。利

用表 2-9 第(7)栏合计数计算如下：

$$|\bar{e}'|=\frac{\sum|(x-\hat{x}')|}{n}=\frac{9.068\%}{12}=0.756\%$$

本例一年 12 个月流通费率的平均数为

$$\bar{x}=\frac{\sum x}{n}=\frac{80.9\%}{12}=6.74\%$$

而 $|\bar{e}'|=0.756\%$，仅为平均流通费率的 11%，表明对过去 12 个月所作的校正预测是相当一致和比较稳定的。$|\bar{e}'|$越低，一致性越高。

经过校正的预测比原来预测的误差下降，预测的准确程度有所提高。根据上述预测平均误差 $\bar{e}$，可对 2012 年 12 月份流通费率的预测值进行校正。

$$\hat{x}_{2012.12}+\bar{e}=8.43\%+0.467\%=8.9\%$$

为了适应经济情况的发展变化，$\bar{e}$ 值可每年调整一次。

思考与练习

1. 定性预测应注意什么问题？
2. 什么是德尔菲法？它有哪些特点？有哪些优缺点？如何选定专家？
3. 德尔菲法集中整理专家意见时，常用的有什么方法？
4. 什么是累计概率中位数法？各种平均误差指标是怎样计算的？对预测值是怎样进行校正的？
5. 已知某公司选定 10 位专家用德尔菲法进行预测，最后一轮征询意见，对明年利润率的估计的累计概率分布如题表 2-1 所示：

题表 2-1

序号＼概率	1%	12.5%	25%	37.5%	50%	62.5%	75%	87.5%	99%
1	8.0	8.1	8.2	8.3	8.4	8.5	8.6	8.7	8.8
2	7.8	8.0	8.2	8.4	8.6	8.8	8.9	9.0	9.1
3	6.0	6.2	6.5	6.7	7.0	7.2	7.5	7.7	8.0
4	6.0	6.5	7.0	7.5	8.0	8.5	8.6	8.7	9.0
5	5.0	5.5	6.0	6.5	7.0	7.5	8.0	8.5	8.9
6	8.0	8.2	8.3	8.4	8.5	8.6	8.8	9.0	9.2
7	6.5	6.7	7.0	7.7	8.0	8.2	8.4	8.6	8.8

续 表

序号＼概率	1%	12.5%	25%	37.5%	50%	62.5%	75%	87.5%	99%
8	7.2	7.6	8.0	8.2	8.4	8.6	8.8	9.0	9.3
9	9.0	9.2	9.3	9.4	9.5	9.6	9.7	9.8	10.0
10	7.5	8.0	8.2	8.4	8.6	8.8	9.0	9.1	9.5

试用累计概率中位数法：① 计算每种概率的不同意见的平均数，用累计概率确定中位数，作为点估计值。② 当要求预测误差不超过1%时的区间估计，及其区间概率。

6. 上题中的某公司过去10次预测的利润率预测值与观察值资料如题表2-2所示：

题表 2-2

预测编号	观察值 x	预测值 $\hat{x}$
1	7.8	8.6
2	7.2	6.8
3	7.0	6.9
4	6.2	6.4
5	6.4	7.0
6	7.8	6.6
7	8.2	6.9
8	7.4	6.3
9	7.7	7.1
10	8.6	8.2

试根据上表资料计算：① 平均绝对误差；② 预测系统误差；③ 校正预测的平均绝对误差；④ 对上题的点估计值进行校正。

3 线性回归分析预测法

在社会经济活动中，任何现象都有其产生的原因，任何原因都会引起一定的结果，这是一般事物运动的规律。这种规律为我们研究现象之间的数量关系提供了依据。例如，港口各项业务活动的发展变化受到多种因素的影响，如港口运输企业的市场需求量要受到进出口货物数量、腹地经济状况、本地 GDP 水平、工人劳动生产率等因素的影响；又如港口吞吐量的变化，要受到经济景气程度、集疏运系统运输能力的大小、集装箱货物占全部运输量的比例等因素变化的影响。要对某种现象进行预测，就必须考虑影响这一现象的有关因素的数量变化，回归分析预测法便是进行这类定量预测的常用方法之一。

回归分析，简单地说，就是在"平均"的意义下，定量地描述各变量之间的数量关系：依据这些数量关系进行预测，就称为"回归分析预测法"。回归分析预测根据某一(或某些)因素的变动来预测某一事物的变动方向和程度，属于因果预测，因而回归分析预测又称为因果关系预测。

回归分析预测法就是从各种现象之间的相互关系出发，通过对与预测对象有关系的现象变动趋势的分析，推算预测对象未来状态数量表现的一种预测方法。

由于篇幅的原因，本章仅就线性回归分析预测法做详细介绍。

3.1 回归分析的基本概念

3.1.1 回归含义

回归分析法起源于生物学的研究。英国生物学家高尔登(Francis Galton)在 19 世纪末研究遗传特性时，发现父亲的身高与儿子的身高之间有较密切的联系。一般来说，父亲高大，儿子也比较高大；父亲矮小，儿子也偏于矮小。但是，大量的研究资料发现，身高有一种向平均数回归的倾向，即身材很高的父亲，儿子比父亲略矮；反之，很矮的父亲，儿子比父亲略高。这种身高倾向于平均数的特性，就称**回归(regression)**。目前，回归分析的理论与应用，均已达到了成熟的阶段，在港口各种经营管理的预测中的应用也很广泛。例如，依据腹地的各种经济指标，找出港口

货物吞吐量与某些经济指标的相关关系，再根据这些相关经济指标建立一元/多元线性回归模型，预测港口吞吐量的未来发展情况。

3.1.2 回归分析与相关分析

1）相关关系的概念

现实世界中，每一事物的运动都是和它周围的事物互相联系、互相影响着的。因此，反映客观事物运动的各种变量之间也就存在着一定的关系。人们通过实践，发现变量之间的关系可以分成两类。

（1）函数关系。它反映现象之间存在着严格的依存关系。在这种关系中，对于某一变量的每一个数值，都有另一个变量的确定的值与之相对应，并且这种关系可以用一个数学表达式反映出来。例如，堆场面积对于集装箱堆存容量的依存关系，船舶航行速度对于航行时间的依存关系。

（2）相关关系。它反映现象之间存在着非严格的、不确定的依存关系。这种依存关系有两个显著的特点：① 现象之间确实存在数量上的客观内在关系。表现在一个现象发生数量上的变化，会影响另一个现象也相应地发生数量上的变化。例如，码头劳动生产率的提高会影响生产成本的降低。② 现象之间数量依存关系不是确定的，具有一定的随机性。表现在给定自变量的一个数值，因变量会有若干个数值和它对应，并且因变量总是遵循一定规律围绕这些数值的平均数上下波动。其原因是影响因变量发生变化的因素不止一个。例如，集装箱码头生产效率的影响因素除了运作机械的数量外，还有劳动力生产效率、堆场面积大小和泊位数量等因素。

2）回归分析与相关分析

回归分析与相关分析均为研究及测度两个或两个以上变量之间关系的方法。相关分析，是研究两个或两个以上随机变量之间相互依存关系的紧密程度。直线相关时用相关系数表示，曲线相关时用相关指数表示，多元相关时用复相关系数表示。回归分析，是研究某一随机变量（因变量）与其他一个或几个普通变量（自变量）之间的数量变动的关系。由回归分析求出的关系式，称为回归模型。

这两种分析的区别：相关分析研究的都是随机变量，并且不分自变量与因变量；回归分析研究的变量要定出自变量与因变量，并且自变量是确定的普遍变量，因变量是随机变量。

这两种分析的联系：它们是研究现象之间相互依存关系的两个不可分割的方面。在实际工作中，一般先进行相关分析，由相关系数或相关指数的大小决定是否需要进行回归分析。而在相关分析的基础上必须拟合回归模型，以便进行推算、预测。

相关分析与回归分析的主要作用：① 通过对数量关系的研究分析，深入认识现

象之间的相互依存关系;② 通过回归模型进行预测和预报;③ 用于补充缺少的资料。

3.1.3 回归模型的种类

回归模型可以从不同的角度进行分类,常用的分类如下所述。

1) 根据回归模型自变量的多少,回归模型可分为一元回归模型和多元回归模型

一元回归模型是根据某一因变量与一个自变量之间的相关关系建立的模型。例如,根据耐用消费品销售量对进口货运量的相关关系建立的回归模型。多元回归模型是根据某一因变量与两个或两个以上自变量之间的相关关系建立的模型。

2) 根据回归模型是否线性,回归模型可分为线性回归模型和非线性回归模型

在线性回归模型中,因变量与自变量的关系是呈直线型的。例如,码头能源(电力、石油等)消耗量与码头装卸货物数量之间的关系。在非线性回归模型中,因变量与自变量的关系是呈曲线形的,如某运输企业的流通费用率与收入额的关系。

3) 根据回归模型是否带虚拟变量,回归模型可分为普通回归模型和带虚拟变量回归模型

普通回归模型的自变量都是数量变量。虚拟变量回归模型的自变量既有数量变量又有品质变量。

此外,根据回归模型是否用滞后的因变量作自变量,回归模型又可分为无自回归现象的回归模型和自回归模型。

3.1.4 回归分析预测的基本问题

所谓回归分析预测,就是依据回归分析描述的变量之间的数量关系对各经济现象进行预测。回归分析预测的基本原理是在回归分析的基础上,根据某一或某些因素的变动来预测某一事物的变动方向和程度,属于因果预测。在回归分析预测过程中,我们一般应关注以下几个基本问题:

1) 变量间相关关系的定性分析

回归分析预测是一种基于定量描述变量之间的数量关系的预测方法。然而,现象之间是否存在着某些数量关系,却不是这种预测方法本身所能确定的。在进行预测前,必须首先根据一定的经济理论、专业知识和实践经验,对变量间是否存在一定的相关性进行分析。只有明确了现象间确实存在某种数量因果关系,才能谈得上定量分析。而不是不加分析地将两个或两个以上的时间序列资料放在一块进行定量分析。比如,将某大宗散货船公司的货运量与同期公司员工癌症患者数进行相关分析,两者之间的相关系数达 0.9 以上,从而得出患癌症与货运量相关的结论。而事实上,船公司货运量的大小与患癌症无任何关系,之所以会得出这样的

结论在于定量分析预测之前忽视了定性分析。

2）变量因果关系的确定

现象之间的数量因果关系，何者为因，何者为果，预测前必须加以明确。在回归分析预测中，把变量称为自变量，它是引起这一现象变化的原因。例如，海铁联运货运量的变化与运输费用、运输时间、服务效率的高低深度相关，因而通常是将海铁联运货运量作为因变量，后三者作为自变量。

3）数学模型的选择

社会经济现象之间的相关关系虽然与确定的函数关系不同，但在大量观察下，仍可以用一些函数式表示它们之间的数量依存关系；变量之间可能的函数关系式很多，在回归分析预测中，必须正确选择数学模型。如线性回归方程 $y=a+bx$，反映的是因变量 y 与自变量 x 之间存在直接相关关系，如果 y 与 x 不存在直接相关关系，则选择这一数学模型预测就会得出错误的结果。

4）回归方程与回归系数的显著性检验

任何两个变量都可以配合成一个回归方程，而这两个变量之间却不一定存在相关关系，即使进行定性分析后，已知两组变量之间存在相关关系，也不能保证一定是线性相关关系。因此，在运用回归方程预测前必须对回归方程和回归系数进行显著性检验。在统计检验方面，通常是：① 用 F 检验法检验回归方程线性关系的显著性；② 用 t 检验法检验每一个自变量与因变量之间是否存在线性相关关系，即检验每一个自变量和系数 b 是否等于零。当检验出现关系显著的结论，才能运用回归方程进行预测。

3.2 一元线性回归预测法

一元回归预测法，也称单因素预测法。有些生产部门运营的发展变化一般都受多种因素的影响，但如果影响预测对象的是一个因素，或诸多因素中有一个因素是基本的、起决定性作用的，那么就可以考虑应用一元回归方程模拟预测对象的发展变化规律，进而估计预测对象未来变化的趋势。一元回归预测一般可以分为一元线性回归预测和一元曲线回归预测两类，前者是最简单的形式。

一元线性回归预测法，是指两个具有线性关系的变量，配合线性回归模型，根据自变量的变动来预测因变量平均发展趋势的方法。

3.2.1 一元线性回归模型

3.2.1.1 回归模型

进行回归分析时，首先需要确定哪个变量是因变量，哪个变量是自变量。在回

归分析中，被预测或被解释的变量称为**因变量**(**dependent variable**)，用 y 表示。用来预测或用来解释因变量的一个或多个变量称为**自变量**(**independent variable**)，用 x 表示。例如，在分析港口城市 GDP 对港口吞吐量的影响时，目的是要预测一定的 GDP 条件下的港口吞吐量是多少。因此，港口吞吐量是被预测的变量，称为因变量，而用来预测港口吞吐量的港口城市 GDP 就是自变量。

在回归分析中，假定自变量 x 是可控制的，而因变量 y 是随机的，但很多情况下并非如此。本章所讨论的回归方法对于自变量是预先固定的和自变量是随机的情况都适用，但固定自变量的情况比较容易描述，因此下面主要讲述固定自变量的回归问题。

对于具有线性关系的两个变量，可以用一个线性方程来表示它们之间的关系。描述因变量 y 如何依赖于自变量 x 和误差项 ε 的方程称为**回归模型**(**regression model**)。只涉及一个自变量的一元线性回归模型可表示为

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{3-1}$$

在一元线性回归模型中，y 是 x 的线性函数 ($\beta_0 + \beta_1 x$ 部分)加上误差项 ε。$\beta_0 + \beta_1 x$ 反映了由于 x 的变化而引起的 y 的线性变化；ε 是被称为误差项的随机变量，反映了除 x 和 y 之间的线性关系之外的随机因素对 y 的影响，是不能由 x 和 y 之间的线性关系所解释的变异性。式中的 β_0 和 β_1 称为模型的参数。

式(3-1)被称为理论回归模型，对这一模型，有以下几个主要假定：

(1) 因变量 y 与自变量 x 之间具有线性关系。

(2) 在重复抽样中，自变量 x 的取值是固定的，即假定 x 是非随机的。

在上述两个假定下，对于任何一个给定的 x 值，y 的取值都对应着一个分布，因此，$E(y) = \beta_0 + \beta_1 x$ 代表一条直线。但由于单个的数据点是从 y 的分布中抽出来的，可能不在这条直线上，因此，必须包含一个误差项 ε 来描述模型的数据点。

(3) 误差项 ε 是一个期望值为 0 的随机变量，即 $E(\varepsilon) = 0$。这意味着在式(3-1)中，由于 β_0 和 β_1 都是常数，所以有 $E(\beta_0) = \beta_0$，$E(\beta_1) = \beta_1$。因此对于一个给定的 x 值，y 的期望值为 $E(y) = \beta_0 + \beta_1 x$。这实际上等于假定模型的形式为一条直线。

(4) 对于所有的 x 值，ε 的方差 σ^2 都相同。这意味着对于一个特定的 x 值，y 的方差也都等于 σ^2。

(5) 误差项 ε 是一个服从正态分布的随机变量，且独立，即 $\varepsilon \sim N(0, \sigma^2)$。独立性意味着对于一个特定的 x 值，它所对应的 ε 与其他 x 值所对应的 ε 不相关。因此，对于一个特定的 x 值，它所对应的 y 值与其他 x 所对应的 y 值也不相关。这表明，在 x 取某个确定值的情况下，y 的变化由误差项 ε 的方差 σ^2 来决定。当 σ^2 较小时，y 的观察值非常靠近直线；当 σ^2 较大时，y 的观察值将偏离直线。由于 σ^2 是常

数，所以 y 的取值不受 x 取值的影响。由于自变量 x 在数据收集前假设是固定的，因此，对于任何一个给定的 x 值，y 都服从期望值为 $\beta_0+\beta_1 x$、方差为 σ^2 的正态分布，且对于不同的 x 都具有相同方差。关于回归模型的假定，如图 3-1 所示。

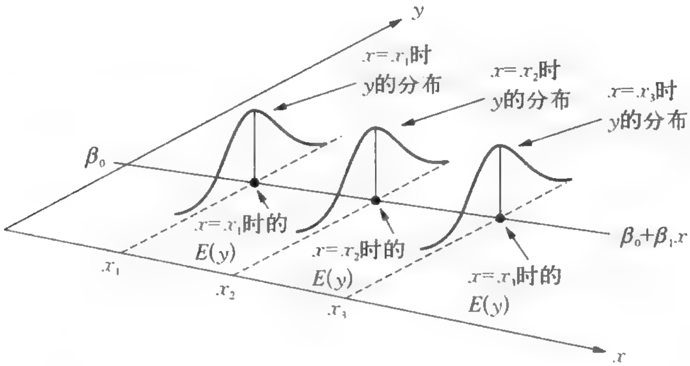


图 3-1 对应不同 x 的 y 和 ε 的分布

从图 3-1 可以看出，$E(y)$ 的值随着 x 的不同而变化，但无论 x 怎样变化，ε 和 y 的概率分布都是正态分布，并且具有相同的方差。

3.2.1.2 回归方程

根据回归模型中的假定，ε 的期望值等于 0，因此 y 的期望值 $E(y)=\beta_0+\beta_1 x$，也就是说，y 的期望值是 x 的线性函数。描述因变量 y 的期望值如何依赖于自变量 x 的方程称为**回归方程（regression equation）**。一元线性回归方程的形式为

$$E(y)=\beta_0+\beta_1 x \tag{3-2}$$

一元线性回归方程的图示是一条直线，因此也称为直线回归方程。其中 β_0 是回归直线在 y 轴上的截距，是当 $x=0$ 时 y 的期望值；β_1 是直线的斜率，它表示当 x 每变动一个单位时，y 的平均变动值。

3.2.1.3 估计的回归方程

如果回归方程中的参数 β_0 和 β_1 已知，对于一个给定的 x 的值，利用式(3-2)就能计算出 y 的期望值。但总体回归参数 β_0 和 β_1 是未知的，必须利用样本数据去估计它们。用样本统计量 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 代替回归方程中的未知参数 β_0 和 β_1，这时就得到了**估计的回归方程（estimated regression equation）**。它是根据样本数据求出的回归方程的估计。

对于一元线性回归，估计的回归方程的形式为

$$\hat{y}=\hat{\beta}_0+\hat{\beta}_1 x \tag{3-3}$$

式中，$\hat{\beta}_0$ 为估计的回归直线在 y 轴上的截距；$\hat{\beta}_1$ 为估计的直线的斜率，表示 x 每变动一个单位时，y 的平均变动值。

3.2.2 参数的最小二乘估计(OLS)

对于第 i 个 x 值，估计的回归方程可表示为

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \tag{3-4}$$

对于 x 和 y 的 n 对观察值，用于描述其关系的直线有多条，究竟用哪条直线来代表两个变量之间的关系，需要有一个明确的原则。我们自然会想到距离各观测点最近的一条直线，用它来代表 x 与 y 之间的关系与实际数据的误差比其他任何直线都小。德国科学家卡尔·高斯(Karl Gauss)提出用最小化图(见图 3-2)中垂直方向的离差平方和来估计参数 β_0 和 β_1，根据这一方法确定模型参数 β_0 和 β_1 的方法称为最小二乘法，也称为**最小平方法(method of least squares)**，它是通过使因变量的观察值 y_i 与估计值 $\hat{y}_i$ 之间的离差平方和达到最小来估计 β_0 和 β_1 的方法。

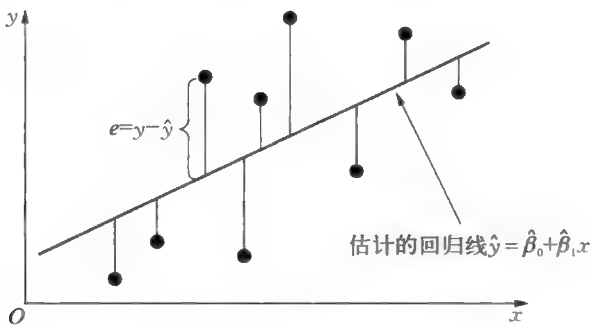


图 3-2 最小二乘法示意图

用最小二乘法拟合的直线具有一些优良的性质。首先，根据最小二乘法得到的回归直线能使离差平方和达到最小，虽然这并不能保证它就是拟合数据的最佳直线(许多其他拟合直线也具有这种性质)，但这毕竟是一条与数据拟合良好的直线应有的性质。其次，由最小二乘法求得的回归直线可知 β_0 和 β_1 的估计量的抽样分布。最后，在某些条件下，β_0 和 β_1 的最小二乘估计量同其他估计量相比，其抽样分布具有较小的标准差。正是基于上述性质，最小二乘法被广泛用于回归模型参数的估计。

根据最小二乘法使

$$\sum (y_i - \hat{y}_i)^2 = \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \tag{3-5}$$

最小。令 $Q=\sum(y_i-\hat{y}_i)^2$，在给定了样本数据后，Q 是 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的函数，且最小值总是存在。根据微积分的极值定理，对 Q 求相应于 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的偏导数，并令其等于 0，便可求出 $\hat{\beta}_0$ 和 $\hat{\beta}_1$，即

$$\begin{cases}\dfrac{\partial Q}{\partial \beta_0}\Big|\beta_0=\hat{\beta}_0=-2\sum(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)^2=0\\[2ex]\dfrac{\partial Q}{\partial \beta_1}\Big|\beta_1=\hat{\beta}_1=-2\sum x_i(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)^2=0\end{cases}\tag{3-6}$$

解上述方程组得

$$\begin{cases}\hat{\beta}_1=\dfrac{n\sum x_iy_i-\sum\limits_{i=1}^{n}x_i\sum\limits_{i=1}^{n}y_i}{n\sum x_i^2-\left(\sum\limits_{i=1}^{n}x_i\right)^2}=\dfrac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2}}\\[2ex]\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}\end{cases}\tag{3-7}$$

由式(3-7)可知，当 $x=\bar{x}$ 时，$y=\bar{y}$，即回归直线 $\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1x_i$ 通过点 $(\bar{x},\bar{y})$，这是回归直线的重要特征之一。

【例 3-1】 某城市一家大型商业银行在很多地区设有分行，其业务主要是进行港航基础设施建设、重点港口项目建设、码头固定资产投资等项目的贷款。近年来，该银行的贷款额平稳增长，但不良贷款额也有较大比例的提高，这给银行业务的发展带来较大压力。为弄清楚不良贷款形成的原因，管理者希望利用银行业务有关数据做些定量分析，以便找出控制不良贷款的方法。表 3-1 就是该银行所属的 25 家分行 2011 年有关业务数据。

表 3-1 某商业银行 2011 年的主要业务数据

	A	B	C	D	E	F
1	分行编号	不良贷款/亿元	各项贷款余额/亿元	本年累计应收贷款/亿元	贷款项目个数/个	本年固定资产投资额/亿元
2	1	0.9	67.3	6.8	5	51.9
3	2	1.1	111.3	19.8	16	90.9
4	3	4.8	173.0	7.7	17	73.7
5	4	3.2	80.8	7.2	10	14.5
6	5	7.8	199.7	16.5	19	63.2

续 表

	A	B	C	D	E	F
7	6	2.7	16.2	2.2	1	2.2
8	7	1.6	107.4	10.7	17	20.2
9	8	12.5	185.4	27.1	18	43.8
10	9	1.0	96.1	1.7	10	55.9
11	10	2.6	72.8	9.1	14	64.3
12	11	0.3	64.2	2.1	11	42.7
13	12	4.0	132.2	11.2	23	76.7
14	13	0.8	58.6	6.0	14	22.8
15	14	3.5	174.6	12.7	26	117.1
16	15	10.2	263.5	15.6	34	146.7
17	16	3.0	79.3	8.9	15	29.9
18	17	0.2	148	0.6	2	42.1
19	18	0.4	73.5	5.9	11	25.3
20	19	1.0	24.7	5.0	4	13.4
21	20	6.8	139.4	7.2	28	64.3
22	21	11.6	368.2	16.8	32	163.9
23	22	1.6	95.7	3.8	10	44.5
24	23	1.2	109.6	10.3	14	67.9
25	24	7.2	196.2	15.8	16	39.7
26	25	3.2	102.2	12.0	10	97.1

管理者想知道，不良贷款是否与贷款余额、累计应收贷款、贷款项目的多少、固定资产投资额等因素相关？如果有关系，它们之间是一种什么关系？关系强度如何？试绘制散点图，并分析不良贷款与贷款余额、累计应收贷款、贷款项目个数、固定资产投资额的关系。图 3－3～图 3－6 为散点图。

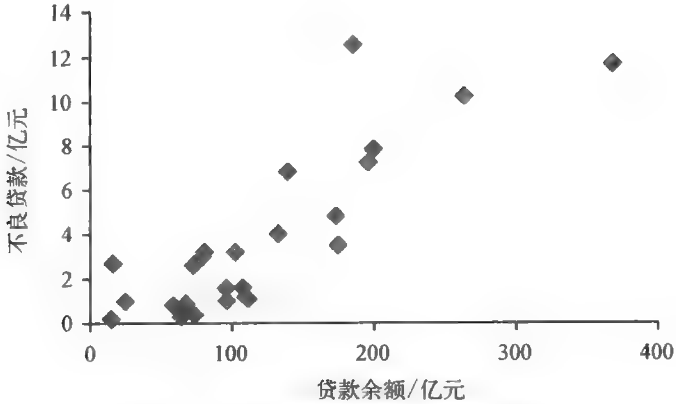


图 3-3　不良贷款与贷款余额的散点图

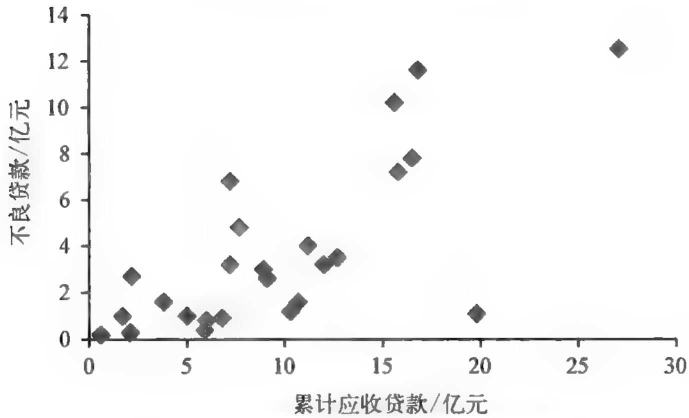


图 3-4　不良贷款与累计应收贷款的散点图

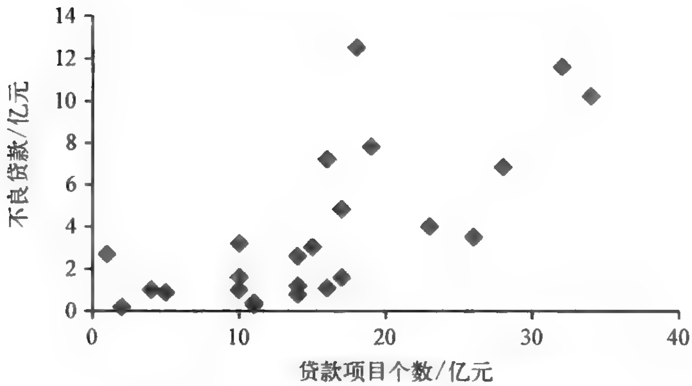


图 3-5　不良贷款与贷款项目个数的散点图

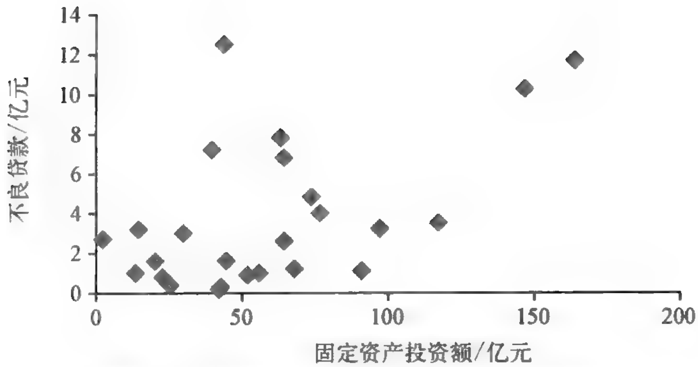


图 3-6　不良贷款与固定资产投资额的散点图

【解】　从散点图可以看出，不良贷款与贷款余额、累计应收贷款、贷款项目个数、固定资产投资额之间都具有一定的线性关系。但从各散点的分布情况看，不良贷款与贷款余额的线性关系比较密切，而与固定资产投资额之间的关系最不密切。

根据式(3-7)，可以求得不良贷款对贷款余额的估计方程。计算如下：

$$\begin{cases}\hat{\beta}_1 = \dfrac{25 \times 17\,080.14 - 3\,006.7 \times 93.2}{25 \times 516\,543.37 - 3\,006.7^2} = 0.037\,895 \\ \hat{\beta}_0 = 3.728 - 0.037\,895 \times 120.268 = -0.829\,5\end{cases}$$

即不良贷款对贷款余额的估计方程为 $\hat{y} = -0.829\,5 + 0.037\,895\,x$。回归系数 $\hat{\beta}_1 = 0.037\,895$ 表示贷款余额每增加 1 亿元，不良贷款平均增加 0.037 895 亿元。在回归分析中，对截距 $\hat{\beta}_0$ 常常不能赋予任何真实意义。例如，在不良贷款与贷款余额的回归中，$\hat{\beta}_0 = -0.829\,5$。如果要解释，它是指当贷款余额为 0 时，不良贷款的平均值为 $-0.829\,5$ 亿元。但是，当贷款余额为 0(没有贷款)时，自然也就不会有不良贷款，而在这里，它的值为一个负数，这就很难解释得通。因此，在回归分析中，对截距 $\hat{\beta}_0$ 通常不做实际意义上的解释。

将 x_i 的各个取值代入上述估计方程，可以得到不良贷款的各个估计值 $\hat{y}_i$。

由图 3-3 可以看出散点图与回归直线的关系。

回归分析中的计算量较大，因此在实际分析中，回归的计算可以用计算机完成。除专门的统计软件外，为大多数人所熟悉的 Excel 也有回归功能。下面，将结合【例 3-1】，说明用 Excel 进行回归的具体步骤。

首先，将不良贷款与贷款余额的数据输入到 Excel 工作表中的 A2：B26 单元格。然后按下列步骤进行操作。

第 1 步：选择【工具】下拉菜单，并选择【数据分析】选项。

第 2 步：在分析工具中选择【回归】，然后单击【确定】按钮。

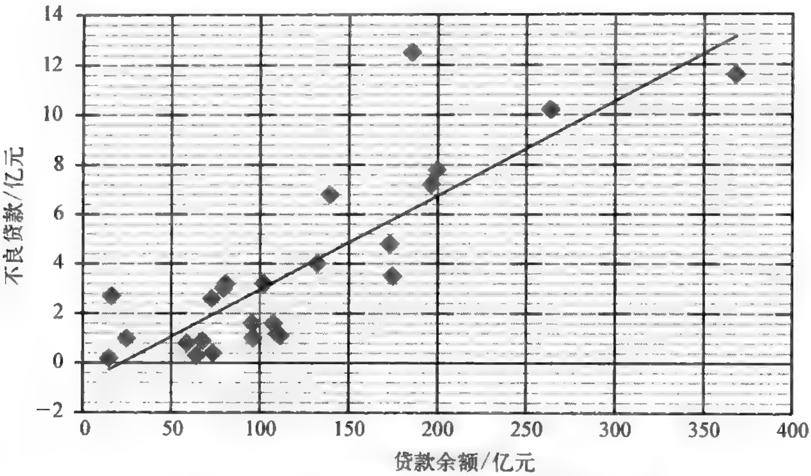


图 3-7 不良贷款对贷款余额的回归直线

第 3 步：当对话框出现时：

在【Y 值输入区域】方框内输入数据区 A2：A26。

在【X 值输入区域】方框内输入数据区 B2：B26。

在【置信度】选项中给出所需的数值(这里常用隐含值 95%)。

在【输出选项】中选择输出区域(这里选新工作表组)。

在【残差】分析选项中选择所需的选项(这里暂时未选)。

结果如图 3-8 所示。

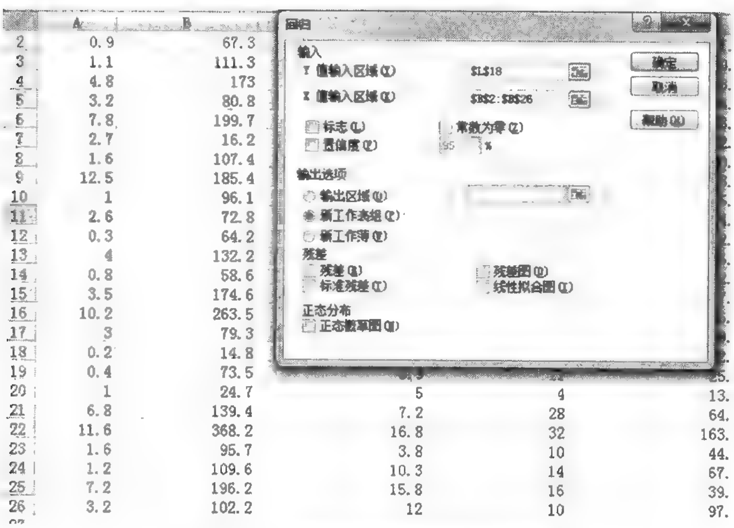


图 3-8 用 Excel 进行回归的步骤

单击【确定】后得到下面的结果，如表 3-2 所示。

表 3-2　Excel 输出的回归分析结果

	A	B	C	D	E	F	G
1	SUMMARY OUTPUT						
2							
3	回归统计						
4	Multiple R	0.843571					
5	R Square	0.711613					
6	Adjusted R Square	0.699074					
7	标准误差	1.979948					
8	观测值	25					
9							
10	方差分析						
11		df	SS	MS	F	Significance F	
12	回归	1	222.48598	222.485979	56.753844	1.18349E-07	
13	残差	23	90.164421	3.920192			
14	总计	24	312.6504				
15							
16		Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%
17	Intercept	-0.829521	0.723043	-1.147263	0.263068	-2.325249	0.666206
18	X Variable 1	0.037895	0.005030	7.533515	0.000000	0.027489	0.048300

Excel 输出的回归结果包括以下几个部分：

第一部分是“回归统计”，这部分给出了回归分析中的一些常用统计量包括**相关系数(multiple R)**、**判定系数(R square)**、**调整的判定系数(adjusted R square)**、标准误差、观测值的个数等。

第二部分是“方差分析”，这部分给出的是回归分析的方差分析表，包括自由度(df)、回归平方和、残差平方和、总平方和(SS)、回归和残差的均方(MS)、检验统计量(F)、**F 检验的显著性水平(significance F)**。“方差分析”部分的主要作用是对回归方程的线性关系进行显著性检验。下面将做详细介绍。

第三部分是回归参数估计的有关内容。包括回归方程的**截距(intercept)**、**斜率(X variable 1)**、截距和斜率的标准误差、用于检验回归系数的 ***t* 统计量(t stat)**和 ***P* 值(P-value)**，以及截距和斜率的置信区间(Lower 95%和 Upper95%)等。

对于本章内容所涉及的一些结果，将在后面陆续介绍。

3.2.3　参数估计量的统计特性

由最小二乘法推导出的回归系数 $\hat{\beta}_1$ 之估计值可以变形为

$$\hat{\beta}_1=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}=\frac{\sum(x_i-\bar{x})y_i-\sum(x_i-\bar{x})y}{\sum(x_i-\bar{x})^2}=\frac{\sum(x_i-\bar{x})y_i}{\sum(x_i-\bar{x})^2}$$

令

$$c_i = \frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2}$$

有

$$\hat{\beta}_1 = \sum c_i y_i \tag{3-8}$$

其中

$$\sum c_i = \frac{\sum (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = 0$$

$$\sum c_i x_i = \sum c_i x_i - \bar{x} \sum c_i = \sum c_i (x_i - \bar{x}) = \sum \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} = 1$$

式(3-8)表明，$\hat{\beta}_1$ 是独立正态变量 $y_1, y_2, \cdots, y_n$ 的线性组合，即估计量 $\hat{\beta}_1$ 具有线性性质。

独立正态变量 $y_1, y_2, \cdots, y_n$ 的线性组合 $\hat{\beta}_1$ 也是正态随机变量，其期望值为

$$\begin{aligned} E(\hat{\beta}_1) &= E\left(\sum c_i y_i\right) = \sum c_i E(y_i) = \sum c_i (\beta_0 + \beta_1 x_i) \\ &= \beta_0 \sum c_i + \hat{\beta}_1 \sum c_i x_i = \beta_1 \end{aligned} \tag{3-9}$$

这表示 $\hat{\beta}_1$ 是 β_1 的无偏估计量。

同理可以证明 $\hat{\beta}_0$ 也是 β_0 的无偏估计量。

进一步研究 $\hat{\beta}_1$ 的方差

$$D(\hat{\beta}_1) = D\left(\sum c_i y_i\right) = \sum c_i^2 D(y_i) = \sum \frac{(x_i - \bar{x})^2}{\left[\sum (x_i - \bar{x})^2\right]^2} \cdot \sigma^2 = \frac{\sigma^2}{\sum (x_i - \bar{x})^2} \tag{3-10}$$

所以

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum (x_i - \bar{x})^2}\right) \tag{3-11}$$

按照上述方法可以求得 $\hat{\beta}_0$ 的方差

$$D(\beta_0) = \sigma^2 \frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2} \tag{3-12}$$

以及 $\hat{\beta}_0$ 与 $\hat{\beta}_1$ 的协方差

$$\begin{aligned}\mathrm{cov}(\hat{\beta}_0,\hat{\beta}_1) &= E\{[\hat{\beta}_0 - E(\hat{\beta}_0)]\cdot[\hat{\beta}_1 - E(\hat{\beta}_1)]\} = E(\hat{\beta}_0-\beta_0)(\hat{\beta}_1-\beta_1) \\ &= -\bar{x}E(\hat{\beta}_1-\beta_1)^2 = -\bar{x}D(\hat{\beta}_1) \qquad (3-13)\end{aligned}$$

可以证明，$\hat{\beta}_0$ 与 $\hat{\beta}_1$ 的方差都具有最小方差性，即采用最小二乘法得到的估计量与用其他方法求出的任何线性无偏估计量相比是最优的（方差最小）。结合式(3-8)和式(3-9)可知，最小二乘估计量 $\hat{\beta}_0$ 、$\hat{\beta}_1$ 具有线性性、无偏性和最小方差性等良好的性质。线性性、无偏性和最小方差性统称 BLUE 性质。满足 BLUE 性质的估计量 $\hat{\beta}_0$ 、$\hat{\beta}_1$ 称为 BLUE 估计量。

3.2.4 回归直线的拟合优度

回归直线 $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ 在一定程度上描述了变量 x 与 y 之间的数量关系，根据这一方程，可依据自变量 x 的取值来估计或预测因变量 y 的取值。但估计或预测的精度如何将取决于回归直线对观测数据的拟合程度。可以想象，如果各观测数据的散点都落在这一直线上，那么这条直线就是对数据的完全拟合，直线充分代表了各个点，此时用 x 来估计 y 是没有误差的。各观察点越是紧密围绕直线，说明直线对观测数据的拟合程度越好，反之则越差。回归直线与各观测点的接近程度称为回归直线对数据的**拟合优度(goodness of fit)**。为说明直线的拟合优度，需要计算判定系数。

3.2.4.1 判定系数 R^2

判定系数是对估计的回归方程拟合优度的度量。为说明它的含义，需要对因变量 y 取值的变差进行研究。因变量 y 的取值是不同的，y 取值的这种波动称为变差。变差的产生来自两个方面：一是由自变量 x 的取值不同造成的；二是除 x 以外的其他因素（如 x 对 y 的非线性影响、测量误差等）的影响。对一个具体的观测值来说，变差的大小可以用实际观测值 y 与其均值 $\bar{y}$ 之差 $(y-\bar{y})$ 来表示。而 n 次观察值的总变差可由这些离差的平方和来表示，称为总平方和，记为 SST，即

$$SST = \sum (y_i - \bar{y})^2 \qquad (3-14)$$

从图 3-9 可以看出，每个观测点的离差都可以分解为

$$y - \bar{y} = (y - \hat{y}) + (\hat{y} - \bar{y}) \qquad (3-15)$$

将式(3-15)两边平方，并对所有 n 个点求和有：

$$\sum (y_i-\bar{y})^2 = \sum (y_i-\hat{y}_i)^2 + \sum (\hat{y}_i-\bar{y})^2 + 2\sum (y_i-\hat{y}_i)(\hat{y}_i-\bar{y}) \qquad (3-16)$$

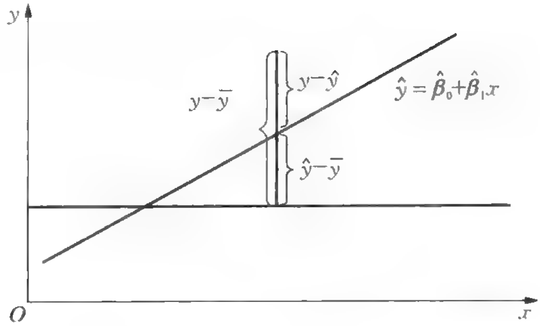


图 3-9 变差分解图

可以证明，$\sum(y_i-\hat{y}_i)(\hat{y}_i-\bar{y})=0$，因此

$$\sum(y_i-\bar{y})^2=\sum(y_i-\hat{y}_i)^2+\sum(\hat{y}_i-\bar{y})^2 \tag{3-17}$$

式(3-17)的左边称为总平方和 SST，它可分解为两部分：其中 $\sum(\hat{y}_i-\bar{y})^2$ 是回归值 $\hat{y}_i$ 与均值 $\bar{y}$ 的离差平方和。根据估计的回归方程，估计值 $\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1x_i$，因此可以把 $(\hat{y}_i-\bar{y})$ 看作由于自变量 x 的变化引起的 y 的变化，而其平方和 $\sum(\hat{y}_i-\bar{y})^2$ 则反映了 y 的总变差中由于 x 与 y 之间的线性关系引起的 y 的变化部分，它是可以由回归直线来解释的 y_i 变差部分，称为回归平方和，记为 SSR。另一部分 $\sum(y_i-\hat{y}_i)^2$ 是各实际观测点与回归值的残差 $(y_i-\hat{y}_i)$ 平方和，它是除了 x 对 y 的线性影响之外的其他因素对 y 变差的作用，是不能由回归直线来解释的 y_i 变差部分，称为残差平方和或误差平方和，记为 SSE。三个平方和的关系为

$$\text{总平方和}(SST)=\text{回归平方和}(SSR)+\text{残差平方和}(SSE) \tag{3-18}$$

从图 3-9 可以直观地看出，回归直线拟合的好坏取决于 SSR 及 SSE 的大小，或者说取决于回归平方和 SSR 占总平方和 SST 的比例 (SSR/SST) 的大小。各观测点越是靠近直线，SSR/SST 则越大，直线拟合得越好。回归平方和占总平方和的比例称为**判定系数(coefficient of determination)**，记为 R^2，其计算公式为

$$R^2=\frac{SSR}{SST}=\frac{\sum(\hat{y}_i-\bar{y})^2}{\sum(y_i-\bar{y})^2}=1-\frac{\sum(y_i-\hat{y}_i)^2}{\sum(y_i-\bar{y})^2} \tag{3-19}$$

判定系数 R^2 测度了回归直线对观测数据的拟合程度。若所有观测点都落在直线上，残差平方和 $SSE=0$，$R^2=1$，拟合是完全的；如果 y 的变化与 x 无关，x 完全无助于解释 y 的变差，此时 $\hat{y}=\bar{y}$，则 $R^2=0$。可见 R^2 的取值范围是[0, 1]。

R^2 越接近于1，表明回归平方和占总平方和的比例越大，回归直线与各观测点越接近，用 x 的变化来解释 y 值变差的部分就越多，回归直线的拟合程度就越好；反之，R^2 越接近于0，回归直线的拟合程度就越差。

3.2.4.2　相关系数 r

相关系数是判定系数的平方根，它是一元线性回归模型中用来衡量两个变量之间相关程度的重要指标。相关系数有两种定义方法：

(1) 根据总变差定义

$$r=\sqrt{\frac{\sum(\hat{y}_i-\bar{y})^2}{\sum(y_i-\bar{y})^2}}=\sqrt{1-\frac{\sum(y_i-\hat{y}_i)^2}{\sum(y_i-\bar{y})^2}}$$

(2) 根据积差法定义，因为

$$\begin{aligned}\frac{\sum(\hat{y}_i-\bar{y})^2}{\sum(y_i-\bar{y})^2}&=\frac{\sum[\hat{\beta}_0+\hat{\beta}_1x_i-\hat{\beta}_0-\hat{\beta}_1\bar{x}]^2}{\sum(y_i-\bar{y})^2}=\frac{\hat{\beta}_1^2\sum(x_i-\bar{x})^2}{\sum(y_i-\bar{y})^2}\\&=\left[\frac{\sum(x_i-\bar{x})^2(y_i-\bar{y})^2}{\sum(x_i-\bar{x})^2}\right]\cdot\frac{\sum(x_i-\bar{x})^2}{\sum(y_i-\bar{y})^2}\\&=\frac{\left[\sum(x_i-\bar{x})(y_i-\bar{y})\right]^2}{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}\end{aligned}$$

所以根据积差法定义的相关系数为

$$r=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2}\sqrt{\sum(y_i-\bar{y})^2}}\tag{3-20}$$

由于根据积差法定义的相关系数不需要先求回归模型的残差平方和，可以直接从样本数据中计算得到，所以在实际工作中用得较为广泛。用积差法计算相关系数计算量比较大，因此，根据平均数的数学性质可将其简化为

$$r=\frac{n\sum x_iy_i-\sum x_i\sum y_i}{\sqrt{n\sum x_i^2-\left(\sum x_i\right)^2}\sqrt{n\sum y_i^2-\left(\sum y_i\right)^2}}\tag{3-21}$$

从上述定义可以看出，相关系数的取值范围为 $-1\leqslant r\leqslant 1$，相关系数为正值表示两变量之间为正相关；相关系数为负值表示两变量之间为负相关。相关系数 r 的绝对值的大小表示相关程度的高低。

i. $|r|=0$ 时，说明回归变差为 0，自变量 x 的变动对总变差毫无影响，这种情况称为零相关。

ii. 当 $|r|=1$ 时，说明回归平方和等于总平方和。总平方和的变化完全由自变量 x 的变化所引起，这种情况称为完全相关。这时自变量 x 与因变量 y 的关系已转化为函数关系。

iii. 当 $0<|r|<1$ 时，说明自变量 x 的变动对总平方和有部分影响，这种情况称为普通相关。其中，r 的绝对值愈大，表示相关程度愈高。一般情况下，当 $|r|\geqslant 0.7$，即 $R^2\geqslant 0.49$ 时，说明自变量 x 的变动对总平方和的影响占一半以上，故称为高度相关；当 $|r|<0.3$，即 $R^2<0.09$ 时，说明自变量 x 的变动对总平方和的影响少于 9%，故称为低度相关；当 $0.3<|r|<0.7$ 时，说明自变量 x 的变动对总平方和的影响程度在 9%～50%之间，故称为中度相关。

比较相关系数 r 与回归系数 $\hat{\beta}_1$ 的计算公式，式(3-7)和式(3-20)，可以发现该两个公式的分母均为正数、分子相同，因此，相关系数 r 与回归系数 $\hat{\beta}_1$ 的正负号是相同的。

【例 3-2】 根据【例 3-1】的数据，计算不良贷款对贷款余额回归的判定系数和相关系数，并解释其意义。

【解】 利用表 3-2，Excel 输出的回归分析结果可知，总平方和 $SST=312.6504$；回归平方和 $SSR=222.4860$；残差平方和 $SSE=90.1644$。根据式(3-19)得

$$R^2=\frac{SSR}{SST}=\frac{222.4860}{312.6504}=0.7116=71.16\%$$

判定系数的实际意义是：在不良贷款取值的变差中，有 71.16%可以由不良贷款与贷款余额之间的线性关系来解释，或者说，在不良贷款取值的变动中，有 71.16%是由贷款余额所决定的。不良贷款取值的差异有 71.16%以上是由贷款余额决定的，可见两者之间有较强的线性关系。

同样，利用表 3-2 Excel 输出的回归分析结果可知，相关系数 $r=0.8435$。

3.2.4.3 估计标准误差 s_e

判定系数可用于度量回归直线的拟合程度，相关系数也可以起到类似的作用。而残差平方和则可以说明实际观测值 y_i 与回归估计值多 $\hat{y}_i$ 之间的差异程度。**估计标准误差(standard error of estimate)**就是度量各实际观测点在直线周围的散布状况的一个统计量，它是均方残差(MSE)的平方根，用 s_e 来表示，其计算公式为

$$s_e=\sqrt{\frac{\sum(y_i-\hat{y}_i)^2}{n-2}}=\sqrt{\frac{SSE}{n-2}}=\sqrt{MSE} \tag{3-22}$$

估计标准误差是对误差项 ε 的标准差 σ 的估计，它可以看作在排除了 x 对 y 的线性影响后，y 随机波动大小的一个估计量。从估计标准误差的实际意义看，它反映了用估计的回归方程预测因变量 y 时预测误差的大小。若各观测点越靠近直线，s_e 越小，回归直线对各观测点的代表性就越好，根据估计的回归方程进行预测也就越准确；若各观测点全部落在直线上，则 $s_e=0$。此时用自变量来预测因变量时是没有误差的。可见 s_e 也从另一个角度说明了回归直线的拟合优度。

从式(3－22)容易看出，回归直线是对 n 个观测点拟合的所有直线中，估计标准误差最小的一条直线，因为回归直线是使 $\sum(y_i-\hat{y}_i)^2$ 为最小时确定的。

【例 3－3】 根据【例 3－1】的有关结果，计算不良贷款对贷款余额回归的估计标准误差，并解释其意义。

【解】 利用表 3－2 Excel 输出的回归分析结果可知，$SSE=90.164\,4$。根据式(3－22)得

$$s_e=\sqrt{\frac{SSE}{n-2}}=\sqrt{\frac{90.164\,4}{25-2}}=1.979\,9(\text{亿元})$$

这就是说，根据贷款余额来估计不良贷款时，平均的估计误差为 1.979 9 亿元。

3.2.5 显著性检验

回归分析的主要目的是根据所建立的估计方程用自变量 x 来估计或预测因变量 y 的取值。当建立了估计方程后，还不能立即进行估计或预测，因为该估计方程是根据样本数据得出的，它是否真实地反映了变量 x 和 y 之间的关系，则需要通过检验后才能证实。

根据样本数据拟合回归方程时，实际上已经假定变量 x 与 y 之间存在着线性关系，即 $y=\beta_0+\beta_1x+\varepsilon$，并假定误差项 ε 是一个服从正态分布的随机变量，且对不同的 x 具有相同的方差。但这些假设是否成立，需要通过检验后才能证实。

回归分析中的显著性检验主要包括两个方面内容：一是线性关系检验；二是回归系数检验。

3.2.5.1 线性关系的检验

线性关系检验是检验自变量 x 和因变量 y 之间的线性关系是否显著，或者说，它们之间能否用一个线性模型 $y=\beta_0+\beta_1x+\varepsilon$ 来表示。为检验两个变量之间的线性关系是否显著，需要构造用于检验的统计量。该统计量的构造是以回归平方和(SSR)和残差平方和(SSE)为基础的。将 SSR 除以其相应的自由度(SSR 的自由度是自变量的个数 k，一元线性回归中自由度为 1)后的结果成为均方回归，记为

MSR；将 SSE 除以其相应的自由度（SSE 的自由度为 $n-k-1$，一元线性回归中自由度为 $n-2$）后的结果称为均方残差，记为 MSE。如果原假设成立（$H_0: \beta_1=0$，两个变量之间的线性关系不显著），则比值 MSR/MSE 的抽样分布服从分子自由度为1、分母自由度为 $n-2$ 的 F 分布，即

$$F=\frac{SSR/1}{SSE/(n-2)}=\frac{MSR}{MSE}\sim F(1,\ n-2) \tag{3-23}$$

所以当原假设 $H_0: \beta_1=0$ 成立时，MSR/MSE 的值应接近1，但如果原假设 $H_0: \beta_1=0$ 不成立，MSR/MSE 的值将变得无穷大。因此，较大的 MSR/MSE 值将导致拒绝原假设 H_0，此时就可以断定变量 x 与 y 之间存在着显著的线性关系。线性关系检验的具体步骤如下。

第1步：提出假设。

$$H_0: \beta_1=0 \quad \text{两个变量之间的线性关系不显著}$$

第2步：计算检验统计量 F。

$$F=\frac{SSR/1}{SSE/(n-2)}=\frac{MSR}{MSE}$$

第3步：做出决策。确定显著性水平 α，并根据分子自由度 $df_1=1$ 和分母自由度 $df_2=n-2$ 查 F 分布表，找到相应的临界值 F_α。若 $F>F_\alpha$，拒绝 H_0，表明两个变量之间的线性关系是显著的；若 $F<F_\alpha$，不拒绝 H_0，没有证据表明两个变量之间的线性关系显著。

【例3-4】 根据【例3-1】的有关结果，检验不良贷款与贷款余额之间线性关系的显著性（$\alpha=0.05$）。

【解】 第1步：提出假设。

$$H_0: \beta_1=0 \quad \text{两个变量之间的线性关系不显著}$$

第2步：计算检验统计量 F。

$$F=\frac{SSR/1}{SSE/(n-2)}=\frac{222.485\,98/1}{90.164\,421/(25-2)}=\frac{222.485\,98}{3.920\,192}=56.753\,84$$

第3步：做出决策。根据显著性水平 $\alpha=0.05$，分子自由度 $df_1=1$ 和分母自由度 $df_2=25-2=23$ 查 F 分布表，找到相应的临界值 $F_\alpha=4.28$。由于 $F>F_\alpha$，拒绝 H_0，表明不良贷款与贷款余额之间的线性关系是显著的。

实际上，在Excel输出的回归结果中，方差分析表部分给出了线性关系显著性检验的全部结果（见表3-2中的方差分析部分）。方差分析表部分除给出了检验

统计的 F 值外，还给出了用于检验的显著性 F，即 Significance F，它就是用于检验的 P 值。将“Significance F”的值与给定的显著性水平 α 的值进行比较，如果 Significance $F<\alpha$，则拒绝原假设 H_0，表明因变量 y 与自变量 x 之间有显著的线性关系；如果 Significance $F>\alpha$，不拒绝原假设 H_0，没有证据表明因变量 y 与自变量 x 之间有显著的线性关系。在表 3 - 2 的输出结果中，Significance $F=1.18349E-07<\alpha=0.05$，这说明不良贷款与贷款余额之间存在显著的线性关系。所得结论与统计量检验相同。

3.2.5.2 回归系数的检验

回归系数的显著性检验是要检验自变量对因变量的影响是否显著。在一元线性回归模型 $y=\beta_0+\beta_1x+\varepsilon$ 中，如果回归系数 $\beta_1=0$，回归线是一条水平线，表明因变量 y 的取值不依赖于自变量 x，即两个变量之间没有线性关系。如果回归系数 $\beta_1\neq 0$，也不能得出两个变量之间存在线性关系的结论，这要看这种关系是否具有统计意义上的显著性。回归系数的显著性检验就是检验回归系数 β_1 是否等于 0。为检验原假设 $H_0:\beta_1=0$ 是否成立，需要构造用于检验的统计量。为此，需要研究回归系数 β_1 的抽样分布。

估计的回归方程 $\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1x_i$ 是根据样本数据计算的。当抽取不同的样本时，就会得出不同的估计方程。实际上，$\hat{\beta}_0$ 和 $\hat{\beta}_1$ 是根据最小二乘法得到的用于估计参数 β_0 和 β_1 的统计量，它们都是随机变量，也都有自己的分布。根据检验的需要，这里只讨论 $\hat{\beta}_1$ 的分布。统计证明，$\hat{\beta}_1$ 服从正态分布，其数学期望为 $E(\hat{\beta}_1)=\beta_1$，标准差为

$$\sigma_{\hat{\beta}_1}=\frac{\sigma}{\sqrt{\sum x_i^2-\frac{1}{n}\left(\sum x_i\right)^2}} \tag{3-24}$$

式中，σ 为误差项 ε 的标准差。

由于 σ 未知，用 σ 的估计量 s_e 代入式(3 - 24)，得到 $\sigma_{\hat{\beta}_1}$ 的估计量，即 β_1 的估计的标准差为

$$s_{\hat{\beta}_1}=\frac{s_e}{\sqrt{\sum x_i^2-\frac{1}{n}\left(\sum x_i\right)^2}} \tag{3-25}$$

这样，就可以构造出用于检验回归系数 β_1 的统计量 t 为

$$t=\frac{\hat{\beta}_1-\beta_1}{s_{\hat{\beta}_1}} \tag{3-26}$$

该统计量服从自由度为 $n-2$ 的 t 分布。如果原假设成立，则 $\beta_1=0$，检验的统计量为

$$t=\frac{\hat{\beta}_1}{s_{\hat{\beta}_1}} \tag{3-27}$$

回归系数显著性检验的具体步骤如下：

第 1 步：提出检验。

$$H_0: \beta_1=0,\ H_1: \beta_1 \neq 0$$

第 2 步：计算检验的统计量 t。

$$t=\frac{\hat{\beta}_1}{s_{\hat{\beta}_1}}$$

第 3 步：做出决策。确定显著性水平 α，并根据自由度 $df=n-2$ 查 t 分布表，找到相应的临界值 $t_{\alpha/2}$。若 $|t|>t_{\alpha/2}$，拒绝 H_0，回归系数等于 0 的可能性小于 α，表明自变量 x 对因变量 y 的影响是显著的。换言之，两个变量之间存在着显著的线性关系；若 $|t|<t_{\alpha/2}$，则不拒绝 H_0，没有证据表明 x 对 y 的影响显著，或者说，两者之间不存在显著的线性关系。

【例 3-5】 根据【例 3-1】的有关结果，检验回归系数的显著性 ($\alpha=0.05$)。

第 1 步：提出假设。

$$H_0: \beta_1=0,\ H_1: \beta_1 \neq 0$$

第 2 步：计算检验的统计量 t。

$$t=\frac{\hat{\beta}_1}{s_{\hat{\beta}_1}}=\frac{0.037\,895}{0.005\,030}=7.533\,797$$

第 3 步：做出决策。根据给定显著性水平 $\alpha=0.05$，自由度 $=n-2=25-2=23$，查 t 分布表，得 $t_{\alpha/2}=t_{0.025}=2.068\,7$。由于 $t=7.533\,797>t_{0.025}=2.068\,7$，拒绝原假设 H_0。这意味着贷款余额是影响不良贷款的一个显著性因素。

在实际应用中，可以直接利用 Excel 输出的参数估计表部分进行检验。表中除了给出检验的统计量，还给出了用于检验的 P 值。检验时可直接将 P 值与给定的显著性水平 α 进行比较。若 $P<\alpha$，则拒绝假设 H_0；若 $P>\alpha$，则不拒绝假设 H_0。在本例中，$P=0.000\,000<\alpha=0.05$，所以拒绝 H_0。

在一元线性回归中，自变量只有一个，上面介绍的 F 检验和 t 检验是等价的；也就是说，如果 $H_0: \beta_1=0$ 被 t 检验拒绝，它也将被 F 检验拒绝。但在多元回归分

析中，这两种检验的意义是不同的。F 检验只是用来检验总体回归关系的显著性，而 t 检验则是检验各个回归系数的显著性。

3.2.5.3 回归分析结果的评价

前面讨论了建立一元线性回归模型的方法。现在的问题是：已经建立的模型是否合适？或者说，这个拟合的模型有多好？要回答这些问题，可从以下几个方面入手：

(1) 所估计的回归系数 $\hat{\beta}_1$ 的符号是否与理论或事先预期相一致。例如，在不良贷款与贷款余额的回归中，贷款余额越多，不良贷款也可能越多；也就是说，回归系数 $\hat{\beta}_1$ 的值应该是正的，在上面建立的回归方程中，得到的回归系数 $\hat{\beta}_1 = 0.037\,895$，为正值。

(2) 如果理论上认为 y 与 x 之间的关系不仅是正的，而且是统计上显著的，那么所建立的回归方程也应该如此。例如，在不良贷款与贷款余额的回归中，两者之间为正的线性关系，而且对回归系数 $\hat{\beta}_1$ 的 t 检验结果表明，两者之间的线性关系是统计上显著的。

(3) 回归模型在多大程度上解释了因变量 y 取值的差异？可以用判定系数 R^2 来回答这一问题。例如，在不良贷款与贷款余额的回归中，得到的 $R^2 = 71.16\%$，解释了在不良贷款变差的 2/3 以上，说明拟合的效果还算不错。

(4) 考察关于误差项 ε 的正态性假定是否成立。因为在对线性关系进行 F 检验和对回归系数进行 t 检验时，都要求误差项 ε 服从正态分布，否则，所用的检验程序将是无效的。检验 ε 正态性的简单方法是画出残差的直方图或正态概率图，该部分内容请参考相关统计学书籍，本书不做详细介绍。

3.2.6 利用回归方程进行预测

3.2.6.1 点估计

利用估计的回归方程，对于 x 的一个特定值 x_0，求出 y 的一个估计值就是点估计。点估计可分为两种：一是平均值的点估计，是对总体参数的估计；二是个别值的点估计，是对因变量的某个具体数值的估计。

平均值的点估计是利用估计的回归方程，对于 x 的一个特定值 x_0，求出 y 的平均值的一个估计值 $E(y_0)$。如在【例 3－1】中，得到的估计的回归方程为 $\hat{y} = -0.829\,5 + 0.037\,895x$，如果要估计贷款余额为 100 亿元，所有分行不良贷款的平均值就是平均值的点估计。根据估计的回归方程，得

$$E(y_0) = -0.829\,5 + 0.037\,895 \times 100 = 2.96\ (\text{亿元})$$

个别值的点估计是利用估计的回归方程，对于 x 一个特定值 x_0，求出 y 的一

个个别值的估计值 $\hat{y}_0$。例如，如果只想知道贷款余额为 72.8 亿元的那个分行(这里是编号为 10 的那个分行)的不良贷款是多少，则属于个别值的点估计。根据估计的回归方程，得

$$\hat{y} = -0.8295 + 0.037895 \times 72.8 = 1.93\ (\text{亿元})$$

这就是说，贷款余额为 72.8 亿元的那个分行的不良贷款估计值为 1.93 亿元。

在点估计条件下，对于同一个 x_0，平均值的点估计和个别值的点估计的结果是一样的，但在区间估计中则有所不同。

3.2.6.2　区间估计

利用估计的回归方程，对于 x 的一个特定值 x_0，求出 y 的一个估计值的区间就是区间估计。区间估计也有两种类型：一是置信区间估计，它是对 x 的一个给定值 x_0，求出 y 的平均值的估计区间，这一区间称为**置信区间(confidence interval)**；二是预测区间估计，它是对 x 的一个给定值 x_0，求出 y 的一个个别值的估计区间，这一区间称为**预测区间(prediction interval)**。

1) y 的平均值的置信区间估计

置信区间估计(confidence interval estimate)是对 x 的一个给定值 x_0，求出 y 的平均值的区间估计。

设 x_0 为自变量 x 的一个特定值或给定值；$E(y_0)$ 为给定 x_0 时因变量 y 的平均值或期望值。当 $x = x_0$ 时，$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$ 为 $E(y_0)$ 的估计值。

一般来说，不能期望估计值 $\hat{y}_0$ 精确地等于 $E(y_0)$。因此，要想用 $\hat{y}_0$ 推断 $E(y_0)$，必须考虑根据估计的回归方程得到的 $\hat{y}_0$ 的方差。对于给定的 x_0，统计学家给出了估计 $\hat{y}_0$ 标准差的公式，用 $s_{\hat{y}_0}$ 表示 $\hat{y}_0$ 标准差的估计量，其计算公式为

$$s_{\hat{y}_0} = s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \tag{3-28}$$

有了 $\hat{y}_0$ 的标准差之后，对于给定的 x_0，$E(y_0)$ 在 $1-\alpha$ 置信水平下的置信区间可表示为

$$\hat{y}_0 \pm t_{\alpha/2} s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \tag{3-29}$$

【例 3-6】　根据【例 3-1】所求得的估计方程，取 $x_0 = 100$，建立不良贷款的 95%的置信区间。

【解】 根据前面的计算结果，已知 $n=25$，$s_e=1.979\ 9$，查表得 $t_{\alpha/2}(n-2)=t_{0.025}(25-2)=2.068\ 7$。

当贷款余额为 100 亿元时，不良贷款的点估计值为

$$E(y_0)=-0.829\ 5+0.037\ 895\times 100=2.96\ (亿元)$$

根据式(3-29)的 $E(y_0)$ 的置信区间为

$$2.96\pm 2.068\ 7\times 1.979\ 9\times\sqrt{\frac{1}{25}+\frac{(100-120.268)^2}{154\ 933.574\ 4}}=2.96\pm 0.845\ 9$$

即 $2.114\ 1\leqslant E(y_0)\leqslant 3.805\ 9$。这就是说，当贷款余额为 100 亿元时，不良贷款的平均值在 2.114 1 亿元～3.805 9 亿元之间。

当 $x_0=\bar{x}$ 时，$\hat{y}_0$ 的标准差的估计量最小，此时有：$s_{\hat{y}_0}=s_e\sqrt{1/n}$。这就是说，当 $x_0=\bar{x}$ 时，估计是最准确的。x_0 偏离 $\bar{x}$ 越远，y 的平均值的置信区间就变得越宽，估计的效果就越不好。

2) y 的个别值的预测区间估计

预测区间估计(prediction interval estimate)是对 x 的一个给定值 x_0，求出 y 的一个个别值的区间估计。

假如不是估计贷款余额为 100 亿元时所有分行的平均不良贷款，而只希望估计贷款余额为 72.8 亿元的那个分行的不良贷款的区间是多少，这个区间则称为预测区间。

为求出预测区间，首先必须知道用于估计的标准差。统计学家已给出了 y 的一个个别估计值 y_0 的标准差的估计量，用 s_{ind} 表示，其计算公式为

$$s_{ind}=s_e\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}} \tag{3-30}$$

因此，对于给定的 x_0，y 的一个个别值 y_0 在 $1-\alpha$ 置信水平下的预测区间可表示为

$$\hat{y}_0\pm t_{\alpha/2}s_e\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}} \tag{3-31}$$

与式(3-29)相比，式(3-31)的根号内多了一个 1。因此，即使是对同一个 x_0，这两个区间的宽度也是不一样的，预测区间要比置信区间宽一些。

【例 3-7】 根据【例 3-1】所求得的估计方程，建立贷款余额为 72.8 亿元的那个分行不良贷款的 95%的预测区间。

【解】 根据前面的计算结果，已知 $n=25$，$s_e=1.979\,9$，查表得 $t_{\alpha/2}(n-2)=t_{0.025}(25-2)=2.068\,7$。

当贷款余额为 72.8 亿元时，不良贷款的点估计值为

$$\hat{y}_0=-0.829\,5+0.037\,895\times 72.8=1.93\text{（亿元）}$$

不良贷款 95%的预测区间为

$$1.93\pm 2.068\,7\times 1.979\,9\times\sqrt{1+\frac{1}{25}+\frac{(72.8-120.268)^2}{154\,933.574\,4}}=1.93\pm 4.206\,6$$

即 $-2.276\,6\leqslant\hat{y}_0\leqslant 6.136\,6$。这就是说，贷款余额为 72.8 亿元的那个分行，其不良贷款的预测区间在 $-2.276\,6$ 亿元～6.136 6 亿元之间。

表 3-3 给出了 25 家分行不良贷款的置信区间和预测区间。

表 3-3　25 家分行不良贷款的置信区间和预测区间

	A	B	C	D	E	F	G	H
1	分行编号	不良贷款 (y)	贷款余额 (x)	预测 Y	置信区间		置信区间	
2					置信下限	置信上限	预测下限	预测上限
3	1	0.9	67.3	1.720 8	0.733 3	2.708 3	−2.496 4	5.938 0
4	2	1.1	111.3	3.388 2	2.563 6	4.212 8	−0.793 8	7.570 2
5	3	4.8	173.0	5.726 3	4.740 1	6.712 4	1.509 4	9.943 1
6	4	3.2	80.8	2.232 4	1.315 9	3.148 9	−1.968 7	6.433 5
7	5	7.8	199.7	6.738 1	5.574 2	7.901 9	2.476 1	11.000 0
8	6	2.7	16.2	−0.215 6	−1.573 7	1.142 4	−4.534 6	4.103 4
9	7	1.6	107.4	3.240 4	2.410 2	4.070 5	−0.942 8	7.423 5
10	8	12.5	185.4	6.196 2	5.132 8	7.259 5	1.960 6	10.413 7
11	9	1.0	96.1	2.812 2	1.955 1	3.669 2	−1.376 4	7.000 7
12	10	2.6	72.8	1.929 2	0.972 5	2.885 9	−2.280 9	6.139 3
13	11	0.3	64.2	1.603 3	0.597 5	2.609 2	−2.618 2	5.824 8
14	12	4.0	132.2	4.180 2	3.351 5	5.008 8	−0.002 7	8.363 0
15	13	0.8	58.6	1.391 1	0.350 4	2.431 9	−2.838 8	5.621 1

续 表

	A	B	C	D	E	F	G	H
16	14	3.5	174.6	5.786 9	4.791 4	6.782 4	1.567 8	10.005 9
17	15	10.2	263.5	9.155 7	7.454 7	10.856 7	4.717 0	13.594 5
18	16	3.0	79.3	2.175 5	1.251 9	3.099 1	−2.027 1	6.378 2
19	17	0.2	14.8	−0.268 7	−1.638 4	1.101 0	−4.591 3	4.054 0
20	18	0.4	73.5	1.955 7	1.002 8	2.908 7	−2.253 5	6.165 0
21	19	1.0	24.7	0.106 5	−1.182 1	1.395 1	−4.191 2	4.404 1
22	20	6.8	139.4	4.453 0	3.609 9	5.296 2	0.267 3	8.638 7
23	21	11.6	368.2	13.123 3	10.416 0	15.835 6	8.210 2	18.036 4
24	22	1.6	95.7	2.797 0	1.938 7	3.655 3	−1.391 8	6.985 8
25	23	1.2	109.6	3.323 7	2.496 9	4.150 5	−0.858 7	7.506 2
26	24	7.2	196.2	6.605 4	5.467 1	7.743 7	2.350 4	10.860 4
27	25	3.2	102.2	3.043 3	2.202 7	3.883 9	−1.141 9	7.228 5

从表 3-3 可以看出，两个区间的宽度不太一样，y 的个别值的预测区间要宽一些。两者的差别表明，估计 y 的平均值比预测 y 的一个特定值或个别值更精确。(请读者想想这是为什么?)同样，当 $x_0=\bar{x}$ 时，预测区间也是最精确的。图 3-10 给出了置信区间和预测区间的示意图。

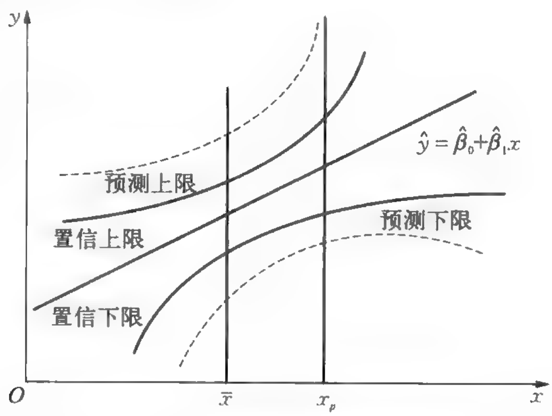


图 3-10 置信区间和预测区间示意图

最后需要注意的是：在利用回归方程进行估计或预测时，不要用样本数据之外的 x 值去预测相对应的 y 值。因为在一元线性回归分析中，总是假定因变量 y 与自变量 x 之间的关系用线性模型表达是正确的。但在实际应用中，它们之间的关系可能是某种曲线。如果 x 的取值范围在 $x_L \sim x_U$ 之间，那么可以利用回归方程对处于 $x_L \sim x_U$ 之间的 x 值来估计 $E(y)$ 和预测 y。但如果用 $x_L \sim x_U$ 之外的 x 值得出的估计值和预测值就会很不理想。

3.3 多元线性回归预测法

3.3.1 多元线性回归模型

在许多实际问题中，影响因变量的因素往往有多个，这种一个因变量同多个自变量的回归问题就是多元回归，当因变量与各自变量之间为线性关系时，称为多元线性回归。多元线性回归分析的原理同一元线性回归基本相同，但在计算上要复杂得多，因此需借助计算机来完成。

3.3.1.1 多元线性回归模型与回归方程

设因变量为 y，k 个自变量分别为 $x_1, x_2, \cdots, x_k$，描述因变量 y 如何依赖于自变量 $x_1, x_2, \cdots, x_k$ 和误差项 ε 的方程称为**多元回归模型(multiple regression model)**。其一般形式可表示为

$$y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \tag{3-32}$$

式中，$\beta_1, \beta_2, \cdots, \beta_k$ 为模型的参数；ε 为误差项。

式(3-32)表明：y 是 $x_1, x_2, \cdots, x_k$ 的线性函数（$\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$ 部分）加上误差项 ε。误差项反映了除 $x_1, x_2, \cdots, x_k$ 与 y 的线性关系之外的随机因素对 y 的影响，是不能由 $x_1, x_2, \cdots, x_k$ 与 y 之间的线性关系所解释的变异性。

与一元线性回归类似，在多元线性回归模型中，对误差项 ε 同样有三个基本的假定：

(1) 误差项 ε 是一个期望值为 0 的随机变量，即 $E(\varepsilon) = 0$。这意味着对于给定 $x_1, x_2, \cdots, x_k$ 的值，y 的期望值为 $E(y) = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$。

(2) 对于自变量 $x_1, x_2, \cdots, x_k$ 的所有值，ε 的方差 σ^2 都相同。

(3) 误差项 ε 是一个服从正态分布的随机变量，且相互独立，即 $\varepsilon \sim N(0, \sigma^2)$。独立性意味着自变量 $x_1, x_2, \cdots, x_k$ 的一组特定值所对应的 ε 与 $x_1, x_2, \cdots, x_k$ 任意一组其他值所对应的 ε 不相关。正态性意味着对于给定的 $x_1, x_2, \cdots, x_k$ 的值，因变量 y 也是一个服从正态分布的随机变量。

根据回归模型的假定有

$$E(y)=\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k \tag{3-33}$$

式(3-32)成为**多元回归方程(multiple regression equation)**，它描述了因变量 y 的期望值与自变量 x_1, x_2, …, x_k 之间的关系。

3.3.1.2　估计的多元回归方程

回归方程中的参数 β_1, β_2, …, β_k 是未知的，需要利用样本数据去估计它们。当用样本统计量 $\hat{\beta}_1$, $\hat{\beta}_2$, …, $\hat{\beta}_k$ 去估计回归方程中的未知参数 β_1, β_2, …, β_k 时，就得到了**估计的多元回归方程(estimated multiple regression equation)**，其一般形式为

$$\hat{y}=\hat{\beta}_1 x_1+\hat{\beta}_2 x_2+\cdots+\hat{\beta}_k x_k \tag{3-34}$$

式中，$\hat{\beta}_1$, $\hat{\beta}_2$, …, $\hat{\beta}_k$ 为参数 β_1, β_2, …, β_k 的估计值；$\hat{y}$ 为因变量 y 的估计值。其中的 $\hat{\beta}_1$, $\hat{\beta}_2$, …, $\hat{\beta}_k$ 称为偏回归系数。$\hat{\beta}_2$ 表示当 x_1, x_3, …, x_k 不变时，x_2 每变动一个单位，因变量 y 的平均变动量；$\hat{\beta}_3$ 表示当 x_1, x_2, …, x_k 不变时，x_3 每变动一个单位，因变量 y 的平均变动量，其余偏回归系数的含义类似。

3.3.1.3　参数的最小二乘估计

根据上述多元线性回归模型，

$$y=\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k+\varepsilon$$

给定变量 y, x_1, x_2, …, x_k 的一组观测值 y_i, x_{1i}, x_{2i}, …, x_{ki}，对应地有

$$y_i=\beta_1 x_{1i}+\beta_2 x_{2i}+\cdots+\beta_k x_{ki}+\varepsilon_i \qquad i=1,\ 2,\ \cdots,\ n \tag{3-35}$$

若取 x_1 的观测值恒等于 1，即对任意 i 有 $x_{1i}=1$，则式(3-35)变为

$$y_i=\beta_1+\beta_2 x_{2i}+\cdots+\beta_k x_{ki}+\varepsilon_i \qquad i=1,\ 2,\ \cdots,\ n$$

即

$$\begin{cases} y_1=\beta_1+\beta_2 x_{21}+\cdots+\beta_k x_{k1}+\varepsilon_1 \\ y_2=\beta_1+\beta_2 x_{22}+\cdots+\beta_k x_{k2}+\varepsilon_2 \\ \cdots\cdots \\ y_n=\beta_1+\beta_2 x_{2n}+\cdots+\beta_k x_{kn}+\varepsilon_n \end{cases}$$

用矩阵形式表示为

$$\boldsymbol{Y}=\boldsymbol{X}\boldsymbol{B}+\boldsymbol{\varepsilon}$$

其中

$$\boldsymbol{Y}=\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}\quad \boldsymbol{X}=\begin{pmatrix}1 & x_{21} & \cdots & x_{k1}\\1 & x_{22} & \cdots & x_{k2}\\\vdots & \vdots & \vdots & \vdots\\1 & x_{2n} & \cdots & x_{kn}\end{pmatrix}\quad \boldsymbol{B}=\begin{pmatrix}\beta_1\\\beta_2\\\vdots\\\beta_k\end{pmatrix}\quad \boldsymbol{\varepsilon}=\begin{pmatrix}\varepsilon_1\\\varepsilon_2\\\vdots\\\varepsilon_n\end{pmatrix}$$

为了估计参数 $\hat{\beta}_1$，$\hat{\beta}_2$，…，$\hat{\beta}_k$，我们仍采用最小平方法。设观察值与模型估计值的残差为 E，则

$$\boldsymbol{E}=\boldsymbol{Y}-\hat{\boldsymbol{Y}}$$

其中

$$\hat{\boldsymbol{Y}}=\boldsymbol{X}\boldsymbol{B}$$

据最小二乘法要求，应有：

$$\boldsymbol{E}'\boldsymbol{E}=(\boldsymbol{Y}-\hat{\boldsymbol{Y}})'(\boldsymbol{Y}-\hat{\boldsymbol{Y}})=\text{最小值} \tag{3-36}$$

即

$$\boldsymbol{E}'\boldsymbol{E}=(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B})'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B})=\text{最小值}$$

由极值原理，根据矩阵求导法则，对其求导并令 $\boldsymbol{B}$ 等于零，则得

$$\frac{\partial \boldsymbol{E}'\boldsymbol{E}}{\partial \boldsymbol{B}}=\frac{\partial(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B})'(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{B})}{\partial \boldsymbol{B}}=\frac{\partial(\boldsymbol{Y}'\boldsymbol{Y}-2\boldsymbol{Y}'\boldsymbol{X}\boldsymbol{B}+\boldsymbol{B}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{B})}{\partial \boldsymbol{B}}$$
$$=-2(\boldsymbol{Y}'\boldsymbol{X}')+2(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{B}=0$$

整理得到回归系数向量的估计值为

$$\hat{\boldsymbol{B}}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \tag{3-37}$$

求解式(3-37)过程非常复杂，需要借助计算机，可直接由 Excel 给出归回结果。

【例 3-8】 根据【例 3-1】的数据，试建立不良贷款(y)与贷款余额(x_2)、累计应收贷款(x_3)、贷款项目个数(x_4)和固定资产投资额(x_5)的线性回归方程，并解释各回归系数的含义。

【解】 由 Excel 输出的多元线性回归结果如表 3-4 所示。

根据表 3-4 的结果，得到不良贷款与贷款余额、累计应收贷款、贷款项目个数和固定资产投资额的多元线性回归方程为

$$\hat{y}=-1.021\,64x_1+0.040\,039x_2+0.148\,034x_3+0.014\,529x_4-0.029\,193x_5$$

因为取 x_1 的观测值恒等于 1，则多元线性回归方程为

$$\hat{y}=-1.021\,64+0.040\,039x_2+0.148\,034x_3+0.014\,529x_4-0.029\,193x_5$$

表 3-4 Excel 输出的回归分析结果

	A	B	C	D	E	F	G
1	SUMMARY OUTPUT						
2							
3	回归统计						
4	Multiple R	0.893087					
5	R Square	0.797604					
6	Adjusted R Squ	0.757125					
7	标准误差	1.778752					
8	观测值	25					
9							
10	方差分析						
11		df	SS	MS	F	Significance F	
12	回归分析	4	249.3712	62.3428	19.704	1.04E-07	
13	残差	20	63.2792	3.16396			
14	总计	24	312.6504				
15							
16		Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%
17	Intercept	-1.02164	0.782372	-1.30582	0.20643	-2.653639	0.61036
18	X Variable 1	0.040039	0.010434	3.837495	0.00103	0.018275	0.061804
19	X Variable 2	0.148034	0.078794	1.878738	0.07494	-0.016328	0.312396
20	X Variable 3	0.014529	0.083033	0.174983	0.86285	-0.158675	0.187733
21	X Variable 4	-0.029193	0.015073	-1.93677	0.06703	-0.60635	0.002249

各回归系数的实际意义为

$\hat{\beta}_2 = 0.040\,039$ 表示，在累计应收贷款、贷款项目个数和固定资产投资额不变的条件下，贷款余额每增加 1 亿元，不良贷款平均增加 0.040 039 亿元。

$\hat{\beta}_3 = 0.148\,034$ 表示，在贷款余额、贷款项目个数和固定资产投资额不变的条件下，累计应收贷款每增加 1 亿元，不良贷款平均增加 0.148 034 亿元。

$\hat{\beta}_4 = 0.014\,529$ 表示，在贷款余额、累计应收贷款和固定资产投资额不变的条件下，贷款项目个数每增加 1 个，不良贷款平均增加 0.014 529 亿元。

$\hat{\beta}_5 = -0.029\,193$ 表示，在贷款余额、累计应收贷款和贷款项目个数不变的条件下，固定资产投资额每增加 1 亿元，不良贷款平均减少 0.029 193 亿元。

3.3.1.4 回归系数的统计性质

1）回归系数 $\hat{B}$ 的数学期望

$$
\begin{aligned}
E(\hat{B}) &= E[(X'X)^{-1}X'Y] \\
&= E[(X'X)^{-1}X'(XB+\varepsilon)] \\
&= E[(X'X)^{-1}X'XB+(X'X)^{-1}X'\varepsilon] \\
&= E(B) \\
&= B
\end{aligned}
$$

可见 $\hat{B}$ 是 B 的无偏估计量。

2）回归系数 $\hat{B}$ 的协方差

$$\mathrm{cov}(\hat{B},\ \hat{B}') = E[(\hat{B}-B)(\overline{B}-B)']$$

因为

$$\begin{aligned}\hat{B}-B &= (X'X)^{-1}X'(XB+\varepsilon)-B \\ &= (X'X)^{-1}X'\varepsilon\end{aligned}$$

故

$$\begin{aligned}\mathrm{cov}(\hat{B},\ \hat{B}') &= E[(X'X)^{-1}X'\varepsilon\varepsilon'X\,(X'X)^{-1}] \\ &= (X'X)^{-1}X'E(\varepsilon\varepsilon')X\,(X'X)^{-1} \\ &= (X'X)^{-1}X'\sigma^2 IX\,(X'X)^{-1} \\ &= (X'X)^{-1}\sigma^2\end{aligned}$$

3.3.2 多元回归方程的拟合优度

3.3.2.1 多重判定系数

与一元回归类似，对多元线性回归方程，需要用多重判定系数来评价其拟合程度。在一元回归中，曾介绍因变量离差平方和的分解方法，对多元回归中因变量离差平方和的分解也一样，同样有

$$SST = SSR + SSE \tag{3-38}$$

式中，$SST = \sum (y_i - \bar{y})^2$ 为总平方和；$SSR = \sum (\hat{y}_i - \bar{y})^2$ 为回归平方和；$SSE = \sum (y_i - \hat{y}_i)^2$ 为残差平方和。

有了这些平方和，可以将多重判定系数定义如下：

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{3-39}$$

多重判定系数(muitiple coefficient of determination)是多元回归中的回归平方和占总平方和的比例，它是度量多元回归方程拟合程度的一个统计量，反映了在因变量 y 的变差中被估计的回归方程所解释的比例。

对于多重判定系数还有一点需要注意：自变量个数的增加将影响到因变量中被估计的回归方程所解释的变差数量。当增加自变量时，会使预测误差变得比较小，从而减少残差平方和 SSE。由于回归平方和 $SSR = SST - SSE$，当 SSE 变小时，SSR 就会变大，从而使 R^2 变大。如果模型中增加一个自变量，即使这个自变量在统计上并不显著，R^2 也会变大。因此，为避免增加自变量而高估 R^2，统计学家提出用样本量 n 和自变量的个数 k 去调整 R^2，计算出**调整的多重判定系数(adjusted multiple coefficient of determination)**，记为 R_a^2，其计算公式为

$$R_a^2 = 1-(1-R^2)\left(\frac{n-1}{n-k-1}\right) \tag{3-40}$$

R_a^2 的解释与 R^2 类似，不同的是：R_a^2 同时考虑了样本量(n)和模型中自变量的个数(k)的影响，这就使得 R_a^2 的值永远小于 R^2，而且 R_a^2 的值不会由于模型中自变量个数的增加而越来越接近 1。因此，在多元回归分析中，通常用调整的多重判定系数。

R^2 的平方根称为多重相关系数，也称为复相差系数，它度量了因变量同 k 个自变量的相关程度。

根据表 3-4 的输出结果得知：**多重判定系数** (R square) $R^2 = 0.797\,604 = 79.760\,4\%$。其实际意义是：在不良贷款取值的变差中，能被不良贷款与贷款余额、累计应收贷款、贷款项目个数和固定资产投资额的多元回归方程所解释的比例为 79.760 4%。

调整的多重判定系数 (adjusted R square) $R_a^2 = 0.757\,125 = 75.712\,5\%$，其意义与 R^2 类似。它表示：在对样本量和模型中自变量的个数进行调整后，在不良贷款取值的变差中，能被不良贷款与贷款余额、累计应收贷款、贷款项目个数和固定资产投资额的多元回归方程所解释的比例为 75.712 5%。

3.3.2.2 估计标准误差

同一元线性回归一样，多元回归中的估计标准误差也是对误差项 ε 的方差 σ^2 的一个估计值，它在衡量多元回归方程的拟合优度方面也起着重要作用。计算公式为

$$s_e = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-m-1}} = \sqrt{\frac{SSE}{n-m-1}} = \sqrt{MSE} \tag{3-41}$$

式中，$m = k-1$ 为除 x_1 以外自变量的个数。

多元回归中对 s_e 的解释与一元回归类似。由于 s_e 所估计的是预测误差的标准差，其含义是根据自变量 x_2，x_3，…，x_k 来预测因变量 y 时的平均预测误差。

根据表 3-4 的输出结果，$s_e = 1.778\,752$（根据公式计算结果也是一样）。其含义是：根据所建立的多元回归方程，用贷款余额、累计应收贷款、贷款项目个数和固定资产投资额来预测不良贷款时，平均的预测误差为 1.778 752 亿元。

3.3.3 显著性检验

在一元线性回归中，线性关系的检验（F 检验）与回归系数的检验（t 检验）是等价的，这一点很容易理解。比如，F 检验表明不良贷款与贷款余额之间有显著的线

性关系，必然也意味着回归系数不会等于 0，因为只有一个自变量。但在多元回归中，这两种检验不再等价。线性关系检验主要是检验因变量同多个自变量的线性关系是否显著，在除 x_1 以外的 $k-1$ 个自变量中，只要有一个自变量与因变量的线性关系显著，F 检验就能通过，但这不一定意味着每个自变量与因变量的关系都显著。回归系数检验则是对每个回归系数分别进行单独的检验，它主要用于检验每个自变量对因变量的影响是否都显著。如果某个自变量没有通过检验，就意味着这个自变量对因变量的影响不显著，也许就没有必要将这个自变量放进回归模型中了。

3.3.3.1 线性关系检验

线性关系检验是检验因变量 y 与 m 个自变量之间的关系是否显著，也称为总体显著性检验。检验的具体步骤如下。

第 1 步：提出假设。

$$H_0: \beta_2 = \cdots = \beta_k = 0$$

$$H_1: \beta_2, \cdots, \beta_k \text{ 至少有一个不等于 } 0$$

第 2 步：计算检验的统计量 $(m = k-1)$

$$F = \frac{SSR/m}{SSE/(n-m-1)} \sim F(m, n-m-1) \tag{3-42}$$

第 3 步：做出统计决策。给定显著性水平 α，根据分子自由度等于 m，分母自由度等于 $n-m-1$，查 F 分布表得 F_α。若 $F > F_\alpha$，则拒绝原假设；若 $F < F_\alpha$ 则不拒绝原假设。

【例 3-9】 根据【例 3-8】建立的回归方程，对回归方程线性关系的显著性进行检验 $(\alpha = 0.05)$。

【解】 第 1 步：提出假设。

$$H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$$

$$H_1: \beta_2, \beta_3, \beta_4, \beta_5 \text{ 至少有一个不等于 } 0$$

第 2 步：计算检验的统计量

$$F = \frac{SSR/m}{SSE/(n-m-1)} = 19.704\,044$$

第 3 步：做出统计决策。给定显著性水平 $\alpha = 0.05$，根据分子自由度等于 4，分母自由度等于 $25-4-1=20$ 查 F 分布表得 $F_{\alpha=0.05}(4, 20) = 2.87$。由于 $F = 19.704 > F_{\alpha=0.05}(4,20) = 2.87$，拒绝原假设 H_0。这意味着不良贷款与贷款余额、

累计应收贷款、贷款项目个数和固定资产投资额之间的线性关系是显著的。

F 检验表明：不良贷款与贷款余额、累计应收贷款、贷款项目个数和固定资产投资额之间的线性关系显著，但这并不意味着不良贷款与每个变量之间的关系都显著，因为 F 检验说明的是总体的显著性。要判断每个自变量对不良贷款的影响是否显著，需要对各回归系数分别进行 t 检验。

3.3.3.2 回归系数检验和推断

在回归方程通过线性关系检验后，就可以对各个回归系数 β_i 有选择地进行一次或多次检验。但究竟要对哪几个回归系数进行检验，通常需要在建立模型之前做出决定。此外，还应对回归系数检验的个数进行限制。

回归系数检验的具体步骤如下：

第 1 步：提出假设。对于任意参数 β_i $(i=2, 3, \cdots, k)$，有

$$H_0: \beta_i = 0$$
$$H_1: \beta_i \neq 0$$

第 2 步：计算检验的统计量 t。

$$t_i = \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}} \sim t(n-m-1) \tag{3-43}$$

式中，$s_{\hat{\beta}_i}$ 是回归系数 $\hat{\beta}_i$ 的抽样分布的标准差，即

$$s_{\hat{\beta}_i} = \frac{s_e}{\sqrt{\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2}} \tag{3-44}$$

第 3 步：做出统计决策。给定显著性水平 $\alpha = 0.05$，根据自由度等于 $n-m-1$ 查 t 分布表，得 $t_{\alpha/2}$ 的值。若 $|t| > t_{\alpha/2}$，拒绝原假设；$|t| < t_{\alpha/2}$ 则不拒绝原假设。

【例 3-10】 根据【例 3-8】建立的回归方程，对回归方程中各回归系数的显著性进行检验 $(\alpha = 0.05)$。

【解】 第 1 步：提出假设。对于任意参数 β_i $(i=2, 3, 4, 5)$，有

$$H_0: \beta_i = 0$$
$$H_1: \beta_i \neq 0$$

第 2 步：计算检验的统计量 t。

$$t_i = \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}} \tag{3-45}$$

根据表 3-4 的结果可知，$t_2=3.837$，$t_3=1.879$，$t_4=0.175$，$t_5=-1.937$。

第 3 步：做出统计决策。给定显著性水平 $\alpha=0.05$。根据自由度等于 $n-m-1=25-4-1$ 查 t 分布表，得 $t_{\alpha/2}=t_{0.025}=2.086$。这里，只有 β_2 通过了检验，其他 3 个自变量都没有通过检验。

这说明在影响不良贷款的 4 个自变量中，只有贷款余额的影响是显著的，而其他 3 个自变量均不显著。这意味着其他 3 个自变量对预测不良贷款的作用已经不大。假定只选一个自变量来预测不良贷款，就应该选贷款余额。

当然，得出上述分析结论还需要有其他证据，因为累计应收贷款、贷款项目个数和固定资产投资额这 3 个变量没有通过检验，也可能是由其他原因造成的。比如，当贷款余额、累计应收贷款、贷款项目个数和固定资产投资额 4 个自变量之间存在高度相关时，就有可能造成某一个或几个回归系数无法通过检验，但这并不一定意味着没通过检验的那些自变量对因变量的影响就不显著。自变量之间的相关所造成的这种问题，在统计上称为多重共线性，有关这一问题将在第 4 节中再做介绍。

除了对回归系数进行检验，还可以求出各回归系数的置信区间。回归系数 β_i 在 $(1-\alpha)\%$ 置信水平下的置信区间为

$$\hat{\beta}_i \pm t_{\alpha/2}(n-m-1)s_{\hat{\beta}_i} \tag{3-46}$$

比如，在表 3-4 中给出的 β_2 的 95%的置信区间为(0.018 275，0.061 804)。这一置信区间的含义是：在累计应收贷款 x_3、贷款项目个数 x_4 和固定资产投资额 x_5 不变的条件下，贷款余额每增加 1 亿元，不良贷款平均增加额在 0.018 275 亿元～0.061 804亿元之间。其他几个回归系数的置信区间的含义类似。

3.3.4 多重共线性

当回归模型中使用两个或两个以上的自变量时，这些自变量往往会提供多余的信息；也就是说，这些自变量之间彼此相关。比如，在【例 3-8】所建立的回归方程中，使用了 4 个自变量，即贷款余额(x_2)、累计应收贷款(x_3)、贷款项目个数(x_4)和固定资产投资额(x_5)。虽然这 4 个自变量对于预测不良贷款都有作用，但由于这 4 个自变量之间本身存在相关关系，在预测中所用的信息就是重复的。从直观上看，贷款余额(x_2)与累计应收贷款(x_3)之间就有较高的相关关系，这两个变量所提供的预测信息就是重复的，或许只用其中的一个自变量就可以了。其他几个变量之间也有类似的相关情况。

3.3.4.1 多重共线性及其所产生的问题

当回归模型中两个或两个以上的自变量彼此相关时，则称回归模型中存在多

重共线性(multicollinearity)。在实际问题中，所使用的自变量之间存在相关是一件很平常的事，但是在回归分析中存在多重共线性时将会产生某些问题。

首先，变量之间高度相关时，可能会使回归的结果造成混乱，甚至会把分析引入歧途。

比如，在【例 3－8】的回归中，根据表 3－4 输出的结果可知，$F=1.03539\times10^{-6}<\alpha=0.05$，这表明不良贷款 y 与贷款余额(x_2)、累计应收贷款(x_3)、贷款项目个数(x_4)和固定资产投资额(x_5)之间的线性回归方程是显著的。但 4 个回归系数中，只有 β_2 通过了检验（$P=0.001028<\alpha=0.05$），而其他 3 个回归系数均未通过检验(P 分别为 0.074 935，0.862 853，0.067 030，均大于 0.05)。

这种检验结果看起来矛盾，但实际上并不矛盾。因为线性关系检验(F 检验)表明回归方程显著时，这只是说，因变量(不良贷款)至少同 4 个自变量中的一个自变量的线性关系是显著的，并非意味着同每个自变量之间的关系都显著。事实上，4 个自变量在预测不良贷款时可能都有贡献(读者可就 4 个自变量分别进行一元回归加以验证)，只不过是一些自变量的贡献与另一些自变量的贡献相互重叠了。

其次，多重共线性可能对参数估计值的正负号产生影响，特别是 β_i 的正负号有可能同预期的正负号相反。比如，从表 3－4 的输出结果可以看出，在 4 个回归系数中，$\hat{\beta}_5=-0.029$，这意味着固定资产投资额增加时，不良贷款是减少的。但实际情况是否如此呢？不一定。如果仅就不良贷款与固定资产投资额做一元回归，得到的估计方程为：$\hat{y}=0.979961+0.046586x$。这表明固定资产投资额每增加 1 亿元，不良贷款平均增加 0.046 586 亿元。这种情况就是由自变量之间的相关所造成的，因为 4 个自变量放在一起产生了多余的信息。因此，当存在多重共线性时，对回归系数的解释将是危险的。

3.3.4.2 多重共线性的判别

检测多重共线性的方法有多种，其中最简单的一种方法是计算模型中各对自变量之间的相关系数，并对各相关系数进行显著性检验。如果有一个或多个相关系数是显著的，就表示模型中所使用的自变量之间相关，因而存在多重共线性问题。

具体来说，如果出现下列情况，暗示存在多重共线性：

(1) 模型中各对自变量之间显著相关。

(2) 当模型的线性关系检验(F 检验)显著时，几乎所有回归系数 β_i 的 t 检验却不显著。

(3) 回归系数的正负号与预期的相反。

下面仍利用【例 3－8】的数据说明上述判别方法的使用。

【例 3-11】 利用【例 3-8】的数据，按上述方法判别所建立的回归方程是否存在多重共线性。

【解】 首先，计算出 4 个自变量之间的相关系数矩阵，由 Excel 输出的结果如表 3-5 所示。

表 3-5 贷款余额、累计应收贷款、贷款项目个数、固定资产投资额之间的相关矩阵

单位：亿元

	A	B	C	D	E
1		各项贷款余额	本年累计应收贷款	贷款项目个数	本年固定资产投资额
2	各项贷款余额	1			
3	本年累计应收贷款	0.678 772	1		
4	贷款项目个数	0.848 416	0.585 831	1	
5	本年固定资产投资额	0.779 702	0.472 431	0.746 646	1

将各相关系数检验的统计量列入，如表 3-6 所示。

表 3-6 各相关系数检验的统计量 单位：亿元

	A	B	C	D	E
1		各项贷款余额	本年累计应收贷款	贷款项目个数	本年固定资产投资额
2	各项贷款余额	1			
3	本年累计应收贷款	4.432 87	1		
4	贷款项目个数	7.686 824	3.466 726	1	
5	本年固定资产投资额	5.971 918	2.570 663	5.382 848	1

查表得 $t_{\alpha/2}(25-2)=2.068\,7$，由于所有的统计量均大于 $t_{\alpha/2}(25-2)=2.068\,7$，所以均拒绝原假设 H_0，说明这 4 个自变量两两之间都有显著的相关关系，这意味着在回归模型中引入 4 个自变量存在多重共线性问题。

其次，由表 3-4 输出的结果可知，回归模型的线性关系是显著的（$F=$

1.035 39E－06 ＜ $\alpha = 0.05$)，而回归系数检验时却有 3 个没有通过 t 检验，这也暗示了模型中存在多重共线性。

最后，固定资产投资额的回归系数为负号(－0.029 193)，这也与预期的不一致。从不良贷款与固定资产投资额的一元回归中得知，固定资产投资额的增加也会使不良贷款增加。

总之，上述三点都表明回归模型中存在多重共线性。

3.3.4.3 多重共线性的处理

一旦发现模型中存在多重共线性问题，就应采取某种解决措施。至于采取什么样的方法来解决，要看多重共线性的严重程度。下面给出多重共线性问题的一些解决办法。

(1) 将一个或多个相关的自变量从模型中剔除，使保留的自变量尽可能不相关。

(2) 如果要在模型中保留所有的自变量，那就应该：① 避免根据 t 统计量对单个参数 β 进行检验。② 对因变量 y 值的推断(估计或预测)限定在自变量样本值的范围内。

【例 3－12】 利用【例 3－8】所建立的回归方程，对多重共线性问题进行处理。

【解】 首先，考虑将一些相关的自变量从模型中剔除。从表 3－5 的相关矩阵中可以看出，贷款余额与贷款项目个数的相关系数最高，而且从定性角度看，贷款余额与累计应收贷款之间也有很强的相关关系。因此将贷款项目个数和累计应收贷款这两个自变量剔除，建立不良贷款(y)与贷款余额(x_2)和固定资产投资额(x_3)的线性模型。由 Excel 输出的结果如表 3－7 所示。

表 3－7 Excel 输出的回归分析结果

	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT								
2									
3	回归统计								
4	Multiple R	0.872379679							
5	R Square	0.761046305							
6	Adjusted R Squ	0.739323242							
7	标准误差	1.84278653							
8	观测值	25							
9									
10	方差分析								
11		df	SS	MS	F	gnificance F			
12	回归分析	2	237.941432	118.970716	35.0340235	1.4503E-07			
13	残差	22	74.7089683	3.3958622					
14	总计	24	312.6504						
15									
16		Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%	下限 95.0%	上限 95.0%
17	Intercept	-0.443423715	0.69686541	-0.6363118	0.53113829	-1.8886341	1.0017867	-1.88863	1.001787
18	X Variable 1	-0.031902732	0.01495416	-2.1333683	0.04429424	-0.0629158	-0.00089	-0.06292	-0.00089
19	X Variable 2	0.050331776	0.00747693	6.73160745	9.1363E-07	0.03482557	0.065838	0.034826	0.065838

从表 3-7 的输出结果可以看出，线性关系和各回归系数在 0.05 的显著性水平下是显著的，多重共线性问题不存在了。

多重共线性问题带来的主要麻烦是对单个回归系数的解释和检验。在求因变量的置信区间和预测区间时一般不会受其影响，但必须保证用于估计或预测的自变量的值是在样本数据的范围之内。因此，如果仅仅是为了估计或预测，可以将所有自变量都保留在模型中。

最后需要提醒的是：在建立多元线性回归模型时，不要试图引入更多的自变量，除非确实有必要。在相关预测研究中，如果收集到的数据是非实验性质的，在某些情况下得到的结果往往并不令人满意，但这不一定是因为选择的模型不合适，而是数据的质量不好，或者是引入的自变量不合适。

3.4 应用回归预测法应注意的问题

随着我国经济的发展、港口业务量的不断增加，人们将不断地借助回归预测法对经济、港口、运输等领域中出现的各种数量关系进行描述、分析、控制和预测。然而，在具体运用时，应注意以下几个问题。

3.4.1 关于定性分析问题

只有在理论上证实经济现象或相关现象指标之间确实存在相关关系性质之后，才能再利用回归预测法来研究和测定各现象相关关系的数量表现。回归分析不能代替对经济现象相互关系的质的分析，只有在质的分析的基础上，才能测定经济现象在数量上的相互关系。这应该成为运用回归预测法的一条基本原则。

因此，我们首先应该进行定性分析，依靠研究人员的理论知识、专业知识、实际经验和分析研究能力，来确定各种变量之间的相关关系及其影响程度。

应该指出，我们必须积极把握事物从量变到质变的“度”的界限。在许多情况下，现象之间只是在一定的范围内才具有相关关系，超出了这个范围，就可能成为荒谬。例如，集装箱码头作业单位成本与码头通过量只有在一定范围内才具有负相关关系，码头通过量超过了一定的限度，单位成本不但不会降低，反而会增加。因此，在运用时要注意它的作用与范围，超过了这个范围去推断或预测，可能会得出错误的结论。

3.4.2 关于回归预测不能任意外推的问题

回归分析的应用，仅仅是限于原来数据所包括的范围内，相关关系只限于 x 和

y 两方面的数值从最小值到最大值的范围；回归分析中的回归关系限于 x 方面原数值从最小值到最大值的范围(以 x 推算 y 时)。所谓外推，是指把相关关系或回归关系用于超出上述范围之外。由于原来资料只提供了一定范围内的数量关系，在此范围以外是否存在同样的关系，尚未得知。如果有进行外推的充分根据和需要，也应十分慎重，而且不能离开原来的范围太远。

统计学已阐明，回归方程实际观察值 y 的偏差 e 不仅与显著性水平 α 有关(α 越小，e 越大)，与 n 有关(n 越大，e 越小)，而且与观察值 x_0 有关。当 x_0 靠近 $\bar{x}$ 时，e 就小；当 x_0 远离 $\bar{x}$ 时，e 就大，即 $e = e(x_0)$。假如分别作函数 $y = \hat{y} - e$ 和 $y = \hat{y} + e$ 的图形(见图 3 - 11)，那么，它们把回归直线夹在中间，两头都是喇叭形的。

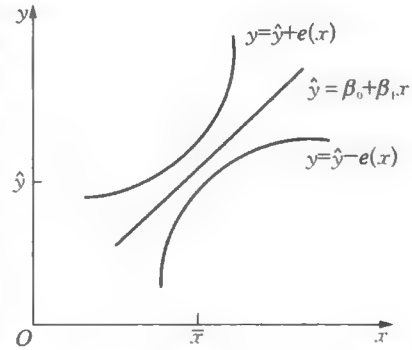


图 3 - 11　小样本预测区间

显而易见，当 x_0 处于 $\bar{x}$ 时，回归直线效果最好。但是，回归直线任意外推，预测远期目标，误差就大，往往会得出不正确的结论。当然，这不能一概而论，但一般来说. 直线回归只适宜做中、短期预测，不宜于做长期预测。

3.4.3 关于对数据资料的要求问题

在利用回归分析进行预测时，还必须注意数据资料的准确性、可比性和独立性问题。

1）关于数据资料的准确性问题

这个问题是容易理解的，只有借以预测的资料是正确可靠的，才能保证分析和预测的可靠性。如果数据是凭经验、拍脑袋估计出来的，就不能得出科学的分析结论。如果由于历史上的某些原因使得某一年度的资料明显不准确(如长三角内河集疏运系统因利用并不充分而导致相关统计数据资料的不充分)，应该按照核实后的数据来计算，不可将错就错。在整理资料过程中，如果发现个别因素缺少某些年度的数字，可采用一定的统计方法(如比例推算法、统计插值法、调查估算法等)予以补齐。如果发现某一年度的数字畸高畸低，可利用数理统计中的控制理论，按照 3δ 原则对该数字进行检验，如与总体平均数的离差超过 3δ，那么该年度各个变量的数值就不能用来分析和推断。

2）关于数据资料的可比性和独立性问题

我们应该保证 y_1，…，y_n 之间的指标数值所包含的内容，指标的口径、范围、

计算方法和计量单位的一致性，并且，各年的指标应当是当年的生产成果。

3) 关于经济现象基本稳定的问题

回归分析是在假定社会经济条件没有发生重大变化、社会经济现象基本稳定的情况下进行的，即假定相关工艺、技术以及国家政策等相对稳定、没有突变的情况下进行的。如果由于某年某个企业发生了重大的技术革命，使生产量成倍地增长，那么，这一年前后的数字就不能合并在一起进行回归预测。

总之，社会经济现象的发展变化是复杂的，这就要求我们在进行回归预测时，必须估计到社会因素的变化，以修正分析的结论。

思考与练习

1. 解释回归模型、回归方程、估计的回归方程的含义。

2. 一元和多元线性回归模型中有哪些基本的假定？

3. 简述参数最小二乘法估计的基本原理。

4. 解释总平方和、回归平方和、残差平方和的含义，并说明它们之间的关系。

5. 简述判定系数、多重判定系数和调整的多重判定系数的含义和作用。

6. 在回归分析中，F 检验和 t 检验各有什么作用？在一元线性回归和多元线性回归中，各有哪些不同？

7. 解释多重共线性的含义。

8. 多重共线性的判定方法有哪些？处理方法有哪些？

9. 在多元线性回归中，选择自变量的方法有哪些？

10. 设 $SSR=36$，$SSE=4$，$n=18$。要求：

(1) 计算判定系数 R^2 并解释其意义。

(2) 计算估计标准误差 s_e 并解释其意义。

11. 一家物流公司的管理人员想研究货物的运送距离和运送时间的关系，为此，他抽出了公司最近 10 辆卡车运货记录的随机样本，得到运送距离(单位：km)和运送时间(单位：天)的数据如题表 3-1 所示。

题表 3-1

运送距离 x	825	215	1 070	550	480	920	1 350	325	670	1 215
运送时间 y	3.5	1.0	4.0	2.0	1.0	3.0	4.5	1.5	3.0	5.0

要求：

(1) 绘制运送距离和运送时间的散点图，判断两者之间的关系形态。

(2) 计算线性相关系数,说明两个变量之间的关系强度。

(3) 利用最小二乘法求出估计的回归方程,并解释回归系数的实际意义。

12. 随机抽取 10 家航空公司,对其最近一年的航班正点率和顾客投诉次数进行了调查,所得数据如题表 3-2 所示。

题表 3-2

航空公司编号	航班正点率	顾客投诉次数(次)
1	81.8	21
2	76.6	58
3	76.6	85
4	75.7	68
5	73.8	74
6	72.2	93
7	71.2	72
8	70.8	122
9	91.4	18
10	68.5	125

要求:

(1) 绘制散点图,说明两者之间的关系形态。

(2) 用航班正点率作自变量,顾客投诉次数作因变量,求出估计的回归方程,并解释回归系数的意义。

(3) 检验回归系数的显著性 ($\alpha = 0.05$)。

(4) 如果航班正点率为 80%,估计顾客投诉次数。

(5) 求航班正点率为 80%时,顾客投诉次数 95%的置信区间和预测区间。

13. 从 $n = 20$ 的样本中得到的有关回归结果是: $SSR = 60$, $SSE = 40$, 要求检验 x 与 y 之间的线性关系是否显著,即检验假设: $H_0: \beta_1 = 0$。

(1) 线性关系检验的统计量 F 值是多少?

(2) 给定显著性水平 $\alpha = 0.05$, F_α 是多少?

(3) 是拒绝原假设还是不拒绝原假设?

(4) 假定 x 与 y 之间是负相关,计算相关系数 r。

(5) 检验 x 与 y 之间的线性关系是否显著。

14. 根据题表 3-3 的数据用 Excel 进行回归,并对回归结果进行讨论,计算 $x_1 = 200$, $x_2 = 7$ 时 y 的预测值。

题表 3-3

y	x_1	x_2
12	174	3
18	281	9
31	189	4
28	202	8
52	149	9
47	188	12
38	215	5
22	150	11
36	167	8
17	135	5

15. 根据下面 Excel 输出的回归结果，说明模型中涉及多少个自变量、多少个观察值？写出回归方程，并根据 F、s_e、R^2 及调整的 R_a^2 的值对模型进行讨论。

	A	B	C	D	E	F
1	SUMMARY OUTPUT					
2						
3	回归统计					
4	Multiple R	0.842 407				
5	R Square	0.709 650				
6	Adjusted R Square	0.630 463				
7	标准误差	109.429 596				
8	观测值	15				
9						
10	方差分析					
11		df	SS	MS	F	Significance F
12	回归分析	3	321 946.801 8	107 315.600 6	8.961 759	0.002 724
13	残差	11	131 723.198 2	11 974.84		
14	总计	14	45 367 0			
15						
16		Coefficients	标准误差	t Stat	P-value	
17	Intercept	657.053 4	167.459 539	3.923 655	0.002 378	
18	X Variable 1	5.710 311	1.791 836	3.186 849	0.008 655	
19	X Variable 2	-0.416 917	0.322 193	-1.293 998	0.222 174	
20	X Variable 3	-3.471 481	1.442 935	-2.405 847	0.034 870	
21						

4 时间序列平滑预测法

时间序列趋势预测法是将预测对象过去的历史资料及数据，按时间顺序加以排列，构成数字系列，并根据其变化动向，预测未来的变化趋势。时间序列趋势预测法属于定量预测法，在港口预测中占据重要地位。本章将重点讲解简易平均法、移动平均法、指数平滑法、差分-指数平滑法、自适应过滤法等的预测方法及预测步骤。

4.1 时间序列分析概述

时间序列趋势预测法是一种经常采用的定量分析方法。它是把某一经济变量的实际观察值按时间先后顺序依次排列，构成一组统计的时间序列，然后应用某种数学方法建立模型，并用此模型来预测该经济变量未来发展变化趋势和变化规律的一种预测技术。

应用时间序列趋势预测法进行预测时，一要注意数据的完整性，二要注意数据资料的可比性，三要保证数据资料的一致性。只有满足了上述三点要求，才能应用这组数据去引申历史，推断未来。

4.1.1 时间序列的因素分析

由于受到多种因素的影响，时间序列数据在不同时期的观察数值存在差异，所呈现出来的变动趋势也不完全一致。这是时间序列数据经常受到一种乃至多种变动因素共同作用的结果。一般情况下，根据各种可能发生变化的因素对时间序列数据变动趋势的影响，按其变动趋势的性质不同，时间序列数据变动趋势分为长期趋势变动、季节变动、循环变动和不规则变动。

1) 长期趋势变动模式

长期趋势变动是指时间序列数据由于受某种根本性因素的影响，使时间序列在较长的时间内朝一定的方向，按线性或非线性变化规律呈上升、水平或下降趋势变化，具体如图 4-1 所示。图 4-1(a)为上升趋势，图 4-1(b)为下降趋势，它们是时间序列在较长时间内表现出的总的趋势，是经济现象本质在数量方面的反映，是时间序列分析和预测的重点。

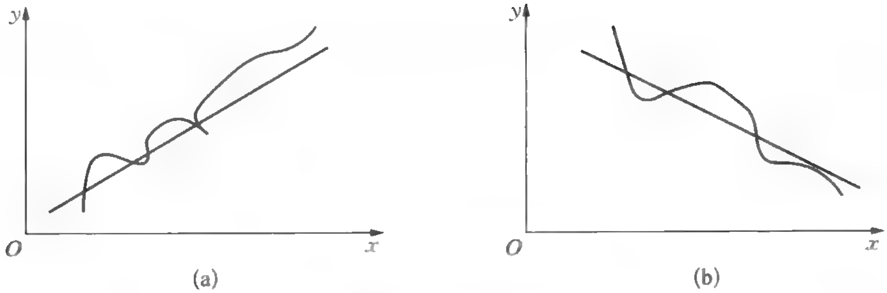


图 4-1 长期趋势示意图

(a) 上升趋势;(b) 下降趋势

2) 循环变动模式

循环变动是以数年(一般不等)为周期的变动。它与长期趋势变动不同。它不是朝着单一方向持续递增(或递减或水平)趋势变化,而是按涨落相间的波浪式起伏变动。它与季节变动趋势也不同。它波动的时间较长,而且变动周期长短不等,短则一两年,长则数年、数十年。

3) 季节变动模式

季节变动是指由于自然条件和社会条件的影响,时间序列数据在一年内随着季节的变化而引起的周期性变动。如随着季节变化,煤炭的需求量也呈周期性变动,因而煤炭的进出口货运量也会呈周期性变化。再如空调器、电风扇、电热器、时装等商品的销售量都会随季节变化而呈周期性波动,则上述商品的进出口贸易总额同样会随季节变化而呈周期性变化。季节变动的周期性比较稳定,一般以年为单位作周期变动。

4) 不规则变动模式

它是指时间序列呈现出的无规律可循的变动,是由随机因素所引起的,如自然灾害、战争、政治暴动、恐怖活动等因素对时间序列的影响。这种不规则的变动对市场的发展影响较大,难以预测。

4.1.2 时间序列的组合形式

时间序列数据是由长期趋势变动、季节变动、循环变动及不规则变动等四种数据类型组成的。其组合方式常见的有以下几种:

1) 加法型

$$Y_t = X_t + C_t + S_t + I_t \tag{4-1}$$

式中,Y_t 为时间为 t 时的序列值;X_t 为时间为 t 时的趋势值;C_t 为时间为 t 时的循

环变动值；S_t 为时间为 t 时的季节变动值；I_t 为时间为 t 时的不规则变动值。

2）乘法型

$$Y_t = X_t C_t S_t I_t \tag{4-2}$$

3）混合型

$$Y_t = T_t C_t + S_t I_t \tag{4-3}$$

或

$$Y_t = T_t C_t S_t + I_t \tag{4-4}$$

对于一个具体的时间序列，应采用哪种组合方式，要根据掌握的数据资料、时间序列的性质及研究的目的等具体情况灵活确定。

4.1.3 时间序列趋势预测法的预测程序

时间序列趋势预测法的预测程序大体包括如下 5 个步骤：

（1）绘制观察期数据的散点图，确定其变化趋势的类型。

（2）对观察期数据加以处理，以消除周期变动、季节性变动和不规则变动因素的影响，仅包括长期趋势变动的影响。

（3）建立数学模型。根据数据处理后的长期趋势变动，结合预测的目的及期限，建立时间序列的预测模型，并对模型进行模拟运算。

（4）修正预测模型。考虑季节变动、循环变动及不规则变动等因素对预测模型的影响并加以修正。

（5）进行预测。采用定量分析与定性分析相结合的方式，对目标变量加以预测，并确定市场未来发展变化的预测值。

4.2 移动平均预测法

移动平均法是将观察期的数据，根据时间序列，逐项推移，依次计算包含一定跨越期的序时平均数，以反映长期趋势的方法。每次移动平均总是在上次移动平均的基础上，去掉一个最远期的数据，增加一个紧挨跨越期后面的新数据，保持跨越期不变，每次只向前移动一步，逐项移动，滚动前移。当时间序列的数值由于受到周期变动和不规则变动的影响，起伏较大，不易显示出发展趋势时，可用移动平均法消除这些因素的影响，分析、预测序列的长期趋势。

移动平均法可以分为简单移动平均法、加权移动平均法和趋势移动平均法等。

4.2.1 简单移动平均

简单移动平均法依照时间顺序，对包含一定项数的历史序列值取算数平均数

作为下一期的预测值。

设时间序列为：y_1，y_2，…，y_m，选取 n 个时期的数据，则第 t 期简单移动平均数的计算公式为

$$M_t=\frac{y_t+y_{t-1}+\cdots+y_{t-n+1}}{n}=\bar{y}_t \tag{4-5}$$

作为 $t+1$ 期的预测值，即以第 t 期移动平均数作为第 $t+1$ 期的预测值，记作：

$$\hat{y}_{t+1}=M_t$$

式中，M_t 为 t 期移动平均数；n 为移动步长。此式表明当 t 向前移动一个时期，就增加一个新近数据，去掉一个远期数据，得到一个新的平均数。

由式(4-5)可得

$$M_{t-1}=\frac{y_{t-1}+y_{t-2}+\cdots+y_{t-n}}{n}$$

$$M_t=\frac{y_t}{n}+\frac{y_{t-1}+y_{t-2}+\cdots+y_{t-n}}{n}-\frac{y_{t-n}}{n}$$

$$M_t=M_{t-1}+\frac{y_t-y_{t-n}}{n} \tag{4-6}$$

即

$$\hat{y}_{t+1}=\hat{y}_t+\frac{y_t-y_{t-n}}{n}$$

这是它的递推公式。当 n 较大时，利用递推公式可以大大减少工作量。式(4-6)表明第 $t+1$ 期的预测值 $\hat{y}_{t+1}$ 实际上是在第 t 期预测值 $\hat{y}_t$ 的基础上加上修正项 $\frac{y_t-y_{t-n}}{n}$。修正项的大小与 n 有关。对于某一时间序列来说，y_t-y_{t-n} 是确定的，n 越大，则 $\left|\frac{y_t-y_{t-n}}{n}\right|$ 越小，这在很大程度上克服了随机干扰的影响，所反映的历史信息越充分，越能显示出趋势的变化，从而修正的作用越大，也就是对数据的平滑能力强，但同时灵敏度降低，使用的成本也增加；反之亦然。所以选择不同的 n，预测结果是不同的。要根据具体情况选择合适的 n 才能达到令人满意的预测精度。选择若干个不同的 n 进行预测，并将预测值与实际值相比较，选择使预测误差最小的 n。

【例 4-1】 某港口 1998—2007 年出口货物总量数据如表 4-1 所示。试用简单移动平均法，预测 2008 年出口货物总量。

【解】 分别取 $n=3$ 和 $n=5$，按预测公式

$$\hat{y}_{t+1}=\frac{y_t+y_{t-1}+y_{t-2}}{3}$$

和

$$\hat{y}_{t+1}=\frac{y_t+y_{t-1}+y_{t-2}+y_{t-3}+y_{t-4}}{5}$$

计算跨越期为 3 年和 5 年移动平均预测值。

利用均方误差公式：

$$\text{当 } n=3 \text{ 时}, MSE=\frac{1}{7}\sum_{t=2001}^{2007}(y_t-\hat{y}_t)^2$$

$$\text{当 } n=5 \text{ 时}, MSE=\frac{1}{5}\sum_{t=2003}^{2007}(y_t-\hat{y}_t)^2$$

其结果列于表 4－1 中。

表 4－1 2008 年某港出口货物总量移动平均法预测值 单位：万吨

年份 t	出口货物总量 y_t	$n=3$		$n=5$	
		理论预测值 $\hat{y}_t$	误差平方	理论预测值 $\hat{y}_t$	误差平方
1998	327.02				
1999	351.47				
2000	378.46				
2001	392.84	352.32	1 641.87		
2002	360.70	374.26	183.87		
2003	367.00	377.33	106.70	362.10	24.01
2004	371.74	373.51	3.13	370.09	2.72
2005	357.72	366.48	76.74	374.15	269.94
2006	364.32	365.49	1.37	370.00	32.26
2007	393.10	364.59	812.62	364.30	829.44
		371.71		370.78	
合计			2 826.30		1 158.37
均方误差			403.76		231.67

由表4-1可知,由于跨越期n的取值不同,同样的历史数据有不同的预测结果,它直接影响预测效果。当n取3时,理论预测值的均方误差为403.76;当n取5时,理论预测值的均方误差就降为231.67,可见对水平型历史数据,跨越期n越大,预测效果就越好。

由表4-1可以看出,实际出口货物总量随机波动较大,经过移动平均法计算后,随机波动显著地减小,即消除了随机干扰,而且求取平均值所用的年数越多,即n越大,修匀的程度也越大,因此波动也越小。但是,在这种情况下,对出口货物总量真实的变化趋势反应也越迟钝。反之,如果n取值越小,对出口货物总量真实的变化趋势反应越灵敏,但修匀性越差,容易把随机干扰作为趋势反映出来。因此,n的选择甚为重要,n应取多大,应该根据具体情况做出抉择。当n等于周期变动的周期时,则可消除周期变动的影响。

简单移动平均法只适合做近期预测,而且是预测目标的发展趋势变化不大的情况。如果目标的发展趋势存在其他变化,采用简单移动平均法就会产生较大的预测偏差和滞后。

4.2.2 加权移动平均法

一般情况下,最近期的经济数据比远期的数据包含了更多未来情况的信息,能更多地反映经济变化的趋势。在简单移动平均公式中,每期数据在平均中的作用是平等的,把各期数据等同看待是不尽合理的,应考虑各期数据的重要性,对近期数据给予较大的权重,对远期的数据给予较小的权重,对于权数因子的确定,完全靠预测者对序列做全面的了解和分析而定,然后再进行移动平均,这就是加权移动平均预测法。

设时间序列为:y_1, y_2, …, y_n, …,加权移动平均公式为

$$M_t = \frac{\omega_1 y_t + \omega_2 y_{t-1} + \cdots + \omega_n y_{t-n+1}}{\sum \omega_i} \tag{4-7}$$

式中,M_t为t期加权移动平均数;ω_i为y_{t-i+1}的加权因子。

利用加权移动平均数来做预测,其预测公式为

$$\hat{y}_{t+1} = M_t \tag{4-8}$$

即以第t期加权移动平均数作为第$t+1$期的预测值。

值得我们注意的是:虽然加权平均法更为科学,能较好地反映近期历史数据对预测值的影响,但它主要适用于呈水平型变动的历史数据,而不适用于趋势型变动的历史数据,否则,就会产生较大的预测误差。

【例 4-2】 某进口煤炭码头的月卸船量记录如表 4-2 所示，适用加权移动平均法预测第 12 个月的卸船量。

表 4-2 加权移动平均预测值 单位：万吨

月份 t	实际月卸船量 y_t	$n=3,\omega_1=1.5,\omega_2=1,\omega_3=0.5$	相对误差(%)	$n=3,\omega_1=1.8,\omega_2=1,\omega_3=0.2$
1	20			
2	30			
3	40			
4	42	33.3	15.95	35.3
5	48	39.3	18.13	40.5
6	50	44.7	10.6	45.5
7	54	48	11.11	48.8
8	60	51.7	13.83	52.3
9	65	56.3	13.38	57.3
10	68	61.5	9.56	62.6
11	73	65.7	10	66.5
12		70		70.8

【解】 取 $\omega_1=1.5$，$\omega_2=1$，$\omega_3=0.5$，按预测公式计算 3 年加权移动平均预测值，其结果列于表 4-2 中。第 12 月份的卸船量的预测值为

$$\hat{y}_{12}=\frac{1.5\times73+1\times68+0.5\times65}{1.5+1+0.5}=70$$

这个预测值偏低，可以修正。其方法是：先计算各年预测值与实际值的相对误差，例如第 4 月份的误差为

$$\frac{42-33.3}{42}\times100\%=15.95\%$$

将相对误差列于表 4-2 中，再计算总的平均相对误差：

$$\left(1-\frac{\sum\hat{y}_t}{\sum y_t}\right)\times100\%=\left(1-\frac{400.5}{460}\right)\times100\%=12.93\%$$

由于总预测值的平均值比实际值低 12.93%，所以可将第 12 月份的预测值修正为

$$\frac{70}{1-12.93\%}=80\text{（万吨）}$$

从表 4-2 中可以看出，选取的权重因子不同，选择的移动步长不同，得到的预测结果是不一样的。如何选择权重因子和移动步长是提高预测精度的关键，往往要借助预测者的经验，选取不同权重方案加以对比，择优选取。

4.2.3 趋势移动平均法

前面介绍的简单移动平均法和加权移动平均法，在时间序列没有明显的趋势变动时，能够准确地反映实际情况。但当时间序列出现直线增加或减少变动趋势时，用简单移动平均法或加权移动平均法来预测就会出现滞后偏差。因此需要进行修正。修正的方法是做二次移动平均，利用移动平均滞后偏差的规律来建立直线趋势的预测模型，这就是趋势移动平均法。

趋势移动平均法对于同时存在直线趋势与周期波动的序列，是一种既能反映趋势变化，又可以有效地分离出周期变化的方法。

二次移动平均法的原理：一次移动平均数总是落后于历史数据，二次移动平均数又总是落后于一次移动平均数，这两个滞后量相等，因此，根据三者间的滞后关系，我们可以先求出一次移动和二次移动平均数间的差值，然后将此差值加到一次移动平均数上，再考虑趋势变动值，就可以得到比较接近实际的预测值，这才是二次移动平均法的基本思想。二次移动平均法的预测模型就是基于这一原理建立的。

一次移动平均数为

$$M_t^{(1)}=\frac{y_t+y_{t-1}+\cdots+y_{t-n+1}}{n}=\bar{y}_t \tag{4-9}$$

在一次移动平均的基础上再进行一次移动平均就是二次移动平均，其计算公式为

$$M_t^{(2)}=\frac{\bar{y}_t+\bar{y}_{t-1}+\cdots+\bar{y}_{t-n+1}}{n}=\bar{y}'_t \tag{4-10}$$

它的递推公式为

$$M_t^{(2)}=M_{t-1}^{(1)}+\frac{M_t^{(1)}-M_{t-n}^{(1)}}{n} \tag{4-11}$$

二次移动平均不是直接将二次移动平均值作为下一期的预测值，而只是用其来求出平滑系数，利用滞后偏差建立线性预测模型，然后再利用模型进行预测。它只适用于存在线性变化趋势的预测问题。其预测模型为

$$\hat{y}_{t+T} = a_t + b_t T \qquad T = 1, 2, \cdots \tag{4-12}$$

式中，$\hat{y}_{t+T}$ 为 $t+T$ 期预测值；t 为当前期数；T 为由 t 至预测期的时期数；a_t、b_t 为平滑系数。

推导出用移动平均值来表示 a_t 和 b_t 的值。

由式(4-12)可知，令 $T=0$ 可得：$a_t = y_t$。

又因为 $\bar{y}'_t - \bar{y}_t = \bar{y}_t - y_t$（一次移动平均数总是落后于历史数据，二次移动平均数又总是落后于一次移动平均数，这两个滞后量相等），即

$$y_t = 2\bar{y}_t - \bar{y}'_t$$

因此
$$a_t = 2\bar{y}_t - \bar{y}'_t \tag{4-13}$$

将式(4-9)变形，可得

$$\begin{aligned}\bar{y}_t &= \frac{y_t + y_{t-1} + \cdots + y_{t-n+1}}{n} \\ &= \frac{y_t + (y_t - b_t) + (y_t - 2b_t) + \cdots + [y_t - (n-1)b_t]}{n} \\ &= y_t - \frac{n-1}{2}b_t \quad \text{（注意：} b_t \text{ 是常数）}\end{aligned}$$

可得
$$y_t - \bar{y}_t = \frac{n-1}{2}b_t$$

所以

$$\bar{y}_t - \bar{y}'_t = \frac{n-1}{2}b_t \tag{4-14}$$

于是，由式(4-13)和式(4-14)可得 a_t 和 b_t 的计算公式：

$$a_t = 2\bar{y}_t - \bar{y}'_t$$
$$b_t = \frac{2(\bar{y}_t - \bar{y}'_t)}{n-1} \tag{4-15}$$

【例 4-3】 1998—2007 年，某港货物吞吐量如表 4-3 所示，用二次移动平均法 ($n=3$) 计算 2003—2007 年该港货物吞吐量的理论预测值，并预测 2008 年的货物吞吐量。

表 4-3 某港货物吞吐量及一次、二次移动平均值计算表 单位：万吨

年份 t	货物吞吐量 y_t	一次移动平均 $M_t^{(1)}$	二次移动平均 $M_t^{(2)}$	$M_t^{(1)}-M_t^{(2)}$	a_t	b_t	预测值 $\hat{y}_{t+T}$
(1)	(2)	(3)	(4)	(5)	(6)=2(3)−(4)	(7)=2×(5)/2	(8)=(6)+(7)
1998	43.97						
1999	43.61						
2000	48.97	45.52					
2001	55.10	49.23					
2002	60.61	54.89	49.88	5.01	59.90	5.01	64.91
2003	63.90	59.87	54.66	5.21	65.08	5.21	70.29
2004	65.65	63.39	59.38	4.01	67.40	4.01	71.41
2005	69.98	66.51	63.26	3.25	69.76	3.25	73.01
2006	69.89	68.51	66.14	2.37	70.88	2.37	73.25
2007	71.49	70.45	68.49	1.96	72.41	1.96	74.37
2008							

【解】 由散点图 4-2 可以看出，该港货物吞吐量基本呈直线上升趋势，可用趋势移动平均法来预测。

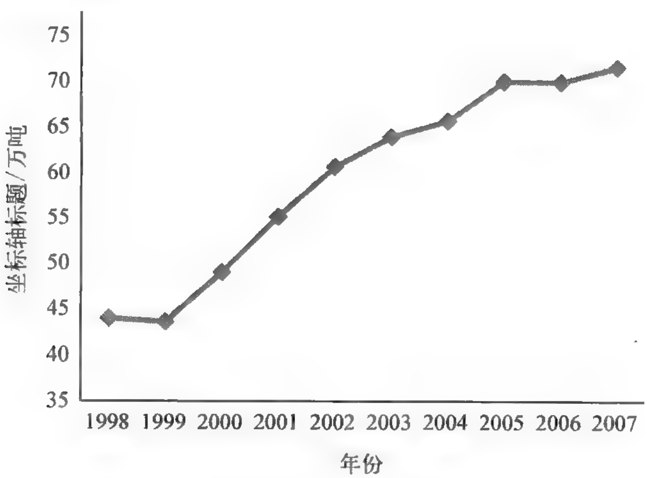


图 4-2 某港 1998—2007 年货物吞吐量

利用式(4－9)计算一次移动平均数,利用式(4－10)计算二次移动平均数,列入表4－3。当 $T=1$ 时,预测期即为 $t+1$ 期, $\hat{y}_{t+T}=a_t+b_tT=a_t+b_t$,将结果列入最后一栏。

预测2009年该港货物吞吐量 $\hat{y}_{12}$,则令 $t=2$,有

$$\hat{y}_{t+2}=a_t+b_t\times 2=72.41+1.96\times 2=76.33\text{(万吨)}$$

比较表中的预测值,可知,利用二次移动法得出的预测值比一次移动平均法得出的预测值更符合实际。可以计算,表中第(3)栏中,2003—2007年5年的均方误差为35.66,而根据表中第(2)栏和第(8)栏的数据,2003—2007年5年的均方误差仅为7.48。这说明对于斜坡型历史数据,利用二次移动平均法进行预测要远比一次移动平均法精确。

4.3 指数平滑法

指数平滑法是一种特殊的加权平均法,权数由对离预测期较近的历史数据到离预测期较远的历史数据按指数规律递减。指数平滑法既不需要存储很多历史数据,又考虑到各期数据的重要性,且使用了全部历史资料。指数平滑法克服了移动平均法存在的两个不足:一是存储数据量大;二是对最近 n 期数据等权看待,而对 $t-n$ 期以前的数据则完全不考虑,这往往不符合实际情况,因此它是移动平均法的改进与发展,应用较为广泛。

指数平滑法,根据平滑次数可分为一次指数平滑法、二次指数平滑法和三次指数平滑法等。

4.3.1 一次指数平滑

4.3.1.1 预测模型

设时间序列为 $y_1, y_2, \cdots, y_t, \cdots$,一次指数平滑公式为

$$s_t^{(1)}=\alpha y_t+(1-\alpha)s_{t-1}^{(1)} \tag{4-16}$$

式中, $s_t^{(1)}$ 为一次指数平滑值; α 为加权系数,且 $0<\alpha<1$; y_t 为第 t 期观测值。则第 $t+1$ 期的预测值即为

$$\hat{y}_{t+1}=s_t^{(1)}$$

一次指数平滑预测法是在移动平均预测法的基础上改进得来的。由式(4－6)可知,移动平均数的递推公式为

$$M_t = M_{t-1} + \frac{y_t - y_{t-n}}{n}$$

因为 y_{t-n} 是参与计算的一个数据点，所以可以认为 M_{t-1} 是 y_{t-n} 的最佳估计值，因而可以用 M_{t-1} 取代 y_{t-n}，则有

$$M_t = M_{t-1} + \frac{y_t - M_{t-1}}{n} = \frac{y_t}{n} + \left(1 - \frac{1}{n}\right)M_{t-1}$$

令 $\alpha = \frac{1}{n}$，以 $s_t^{(1)}$ 代替 M_t，则得式(4-15)

$$s_t^{(1)} = \alpha y_t + (1-\alpha)s_{t-1}^{(1)}$$

指数平滑预测对不同时期的观察值赋予了不同的权值，把式(4-16)递推展开来进一步理解指数平滑的实质：

$$\begin{aligned} s_t^{(1)} &= \alpha y_t + (1-\alpha)[\alpha y_t - (1-\alpha)s_{t-2}^{(1)}] \\ &= \alpha y_t + \alpha(1-\alpha)y_{t-1} + (1-\alpha)^2 s_{t-2}^{(1)} \\ &\cdots\cdots \\ &= \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots + (1-\alpha)^t s_0^{(1)} \\ &= \alpha\sum_{j=0}^{t-1}(1-\alpha)^j y_{t-j} + (1-\alpha)^t s_0^{(1)} \end{aligned} \tag{4-17}$$

由于 $0 < \alpha < 1$，当 $t \to \infty$ 时，$(1-\alpha)^t \to 0$，于是式(4-17)变为

$$s_t^{(1)} = \alpha\sum_{j=0}^{\infty}(1-\alpha)^j y_{t-j} \tag{4-18}$$

可知，第 t 期的指数平滑值是全部观察值的加权线性组合。加权系数分别为 α，$\alpha(1-\alpha)$，$\alpha(1-\alpha)^2$，$\cdots$，$\alpha(1-\alpha)^{t-1}$，是按几何级数衰减，越靠近近期的观察值，权数越大；越远的观察值，权数越小。

加权系数 α，$\alpha(1-\alpha)$，$\alpha(1-\alpha)^2$，$\cdots$，$\alpha(1-\alpha)^{t-1}$ 呈等比级数，公比为 $1-\alpha$，在实际预测中 $0 < \alpha < 1$，则权数之和为

$$s = \frac{\alpha}{1-(1-\alpha)} = 1$$

由于加权系数符合指数规律，又具有平滑数据的功能，故称为指数平滑。以这种平滑值进行预测，就是一次指数平滑法。它的预测模型为

$$\hat{y}_{t+1} = s_t^{(1)}$$

即

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t \tag{4-19}$$

也就是以第 t 期指数平滑值作为 $t+1$ 期预测值。

4.3.1.2　加权系数的选择

在指数平滑法中，加权系数的选择直接影响着预测效果。α 取值是否恰当对预测精度有着关键的决定性作用。由式(4-19)可以看出，α 的大小规定了第 t 期观察值 y_t 在预测下一期中的重要程度。α 值越大，则 y_t 在预测结果中所占的比重就越大，修正幅度越大，对时间序列的近期变化反应越灵敏，但同时也越容易受随机干扰的影响，风险越大。

若选取 $\alpha=0$，则 $\hat{y}_{t+1}=\hat{y}_t$，即下期预测值就等于本期预测值，而完全忽略时间序列的变化；若选取 $\alpha=1$，则 $\hat{y}_{t+1}=\hat{y}_t$，即下期预测值就等于本期观察值，完全不考虑 t 期之前观察值的影响。这两种极端情况很难做出正确的预测。因此，α 值应根据时间序列的具体性质在 0～1 之间进行选择。具体如何选择一般可遵循下列原则：

(1) 如果时间序列波动不大，比较平稳，则 α 应取小一点(一般为 0.1～0.4)，使较早的观察值亦能充分反映于指数平滑中，以减少修正幅度，使预测模型能包含较长时间序列的信息。

(2) 如果时间序列具有迅速且明显的变动倾向，则 α 应取大一点(一般为 0.6～0.8)，使预测模型灵敏度高些，以便迅速跟上数据的变化。

(3) 如果对初始值的正确性有疑问时，应取较大的 α 值，以便扩大近期数据的作用，迅速减少初始值的影响。

(4) 如果多项式模型中仅有某一段时间的数据为较优估计值，则应取较大的 α 值，以减少较早数据的影响。

(5) 如果时间序列虽然具有不规则变动，但长期趋势接近某一稳定常数时，则应取较小的 α 值(一般为 0.05～0.20)，使各观察值在现时指数平滑值中具有大小接近的权值。

在实用上，类似移动平均法，多取几个 α 值进行试算，选取使预测误差的平方和 $Q(\alpha)$ 达到最小的 α：

$$Q(\alpha) = \sum_{t=1}^{n}(y_t - \hat{y}_t)^2 \tag{4-20}$$

如何才能求出最优的 α 呢？一般常用穷举法，即将 α 值离散化，规定一定的步长，在(0, 1)范围内以步长为间隔连续地取值并计算预测误差的平方和 $Q(\alpha)$，找出使 $Q(\alpha)$ 最小的那个 α，即最优的 α。

4.3.1.3 初始值的确定

指数平滑法在进行预测时要求知道前一期的指数平滑预测值，所以必然涉及第一个预测值即初始值如何确定的问题。当时间序列的数据较多，比如在 20 个以上时，初始值对以后的预测值影响很小，可选用第一期数据为初始值。如果时间序列的数据较少，在 20 个以下时，初始值对以后的预测值影响很大。这时，就必须认真研究如何正确确定初始值。有以下方法可供参考：

(1) 若样本序列较长，则可以把已有数据或其中某一部分的算术平均值或加权平均值作为初始值。

$$s_0 = \frac{y_1 + y_2 + \cdots + y_n}{n}$$

或

$$s_0 = \frac{w_1 y_1 + w_2 y_2 + \cdots + w_n y_n}{n}$$

式中，$$w_1 + w_2 + \cdots + w_n = n。$$

(2) 若样本序列不长也不短，可取 $s_0 = y_1$。

(3) 当样本序列的长度 n 充分大时，初始预测值可以任意确定。

如果没有历史数据时，可以采用专家评估法进行评估来确定初始值。预测时可以根据有关过程的先验知识和根据其他类似过程进行类推。

【例 4-4】 2002—2007 年某港口城市平均家庭总收入如表 4-4 第(2)栏所示，试用一次指数平滑法(α 分别取 0.4 和 0.8)，计算 2002—2007 年的理论预测值，并预测 2008 年该港平均家庭总收入。计算 α 分别取 0.4 和 0.8 时的均方误差，来比较预测效果。

表 4-4 一次指数平滑预测法计算表 单位：万美元

年 份	t	家庭总收入 y_t	$\alpha=0.4$		$\alpha=0.8$	
			预测值 $\hat{y}_t$	误差平方	预测值 $\hat{y}_t$	误差平方
(1)		(2)	(3)	(4) = $[(2)-(3)]^2$	(5)	(6) = $[(2)-(5)]^2$
2002	1	15.79	15.79	0	15.79	0
2003	2	16.37	15.79	0.34	15.79	0.34
2004	3	17.23	16.02	1.46	16.25	0.96
2005	4	17.73	16.50	1.51	17.03	0.49
2006	5	21.59	17.00	21.07	17.59	16.00
2007	6	17.17	18.84	2.79	20.79	13.10
2008	7		18.17		17.89	
合计				27.17		30.89
均方误差				4.53		5.15

【解】 将 2002 年该市的平均家庭总收入作为初始预测值

$$s_0^{(1)} = y_1$$

根据预测模型公式：

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t$$

可得第(3)栏、第(5)栏的预测值。如当 $\alpha=0.4$ 时，2006 年的理论预测值为

$$\hat{y}_5 = \alpha y_4 + (1-\alpha)\hat{y}_4 = 0.4 \times 17.73 + (1-0.4) \times 16.50 = 17.0$$

从表 4-4 中可以看出，$\alpha = 0.4$ 和 $\alpha = 0.8$ 时，预测值是很不相同的。究竟 α 取何值为好，可通过计算它们的均方误差 MSE，选取使 MSE 较小的那个 α 值。

由表中计算的结果可知：$\alpha = 0.4$ 时的均方误差较小，可知 α 取 0.4 时预测结果较好。故选取 $\alpha = 0.4$，预测 2008 年该市平均家庭总收入产量为

$$\hat{y}_7 = a y_6 + (1-a)\hat{y}_6 = 0.4 \times 17.17 + (1-0.4) \times 18.84 = 18.17 \text{ 万美元}$$

4.3.2 二次指数平滑

一次指数平滑法虽然克服了移动平均法的两个缺点，它不需要大量的历史数据，计算量小，只要确定好合适的平滑系数 α 和初始值就可以进行预测，但当时间序列的变动出现直线趋势时，用一次指数平滑法进行预测，仍存在着明显的滞后偏差。因此，也必须加以修正。修正的方法与趋势移动平滑法相同，即对一次指数平滑后的序列数据再做一次指数平滑，但并不是直接将二次指数平滑值作为预测值，而是利用其来求出方程参数，利用滞后偏差的规律来建立直线趋势模型。这就是二次指数平滑法。其计算公式为

$$\begin{aligned} s_t^{(1)} &= \alpha y_t + (1-\alpha)s_{t-1}^{(1)} \\ s_t^{(2)} &= \alpha s_t^{(1)} + (1-\alpha)s_{t-1}^{(2)} \end{aligned} \tag{4-21}$$

式中，$s_t^{(1)}$ 为一次指数平滑值；$s_{t-1}^{(2)}$ 为二次指数平滑值。

当时间序列 $\{y_t\}$ 从某时期开始具有直线趋势时，类似趋势移动平均法，可用以下直线趋势模型来预测：

$$\hat{y}_{t+T} = a_t + b_t T \quad T = 1, 2, \cdots \tag{4-22}$$

$$a_t = 2s_t^{(1)} - s_t^{(2)}$$

$$b_t = \frac{\alpha}{1-\alpha}(s_t^{(1)} - s_t^{(2)}) \tag{4-23}$$

下面，我们用增量分析方法来证明式(4－23)。

由式(4－18)可知：

$$s_t^{(1)} = \alpha \sum_{j=0}^{\infty} (1-\alpha)^j y_{t-j}$$

同理

$$s_t^{(2)} = \alpha s_t^{(1)} + (1-\alpha) s_{t-1}^{(2)} = \alpha \sum_{j=0}^{\infty} (1-\alpha)^j s_{t-j}^{(1)}$$

而

$$s_{t-j}^{(1)} = \alpha y_{t-j} + (1-\alpha) s_{t-j-1}^{(1)} = \alpha \sum_{k=0}^{\infty} (1-\alpha)^k y_{t-j-k}$$

所以

$$\begin{aligned} E(s_t^{(1)}) &= \alpha \sum_{j=0}^{\infty} (1-\alpha)^j E(y_{t-j}) \\ &= \alpha \sum_{j=0}^{\infty} (1-\alpha)^j (a_t - b_t j) \\ &= a_t - \frac{1-\alpha}{\alpha} b_t \end{aligned} \tag{4-24}$$

其中

$$\alpha \sum_{j=0}^{\infty} (1-\alpha)^j = 1$$

$$\alpha \sum_{j=0}^{\infty} j \cdot (1-\alpha)^j = \frac{1-\alpha}{\alpha}$$

$$\begin{aligned} E(s_{t-j}^{(1)}) &= \alpha \sum_{k=0}^{\infty} (1-\alpha)^k E(y_{t-j-k}) \\ &= \alpha \sum_{k=0}^{\infty} (1-\alpha)^k [a_t - b_t (j+k)] \\ &= a_t - b_t j - \frac{1-\alpha}{\alpha} b_t \end{aligned}$$

$$\begin{aligned} E(s_{t-j}^{(2)}) &= \alpha \sum_{j=0}^{\infty} (1-\alpha)^j E(s_{t-j}^{(1)}) \\ &= \alpha \sum_{j=0}^{\infty} (1-\alpha)^j \left(a_t - b_t j - \frac{1-\alpha}{\alpha} b_t\right) \\ &= a_t - \frac{2(1-\alpha)}{\alpha} b_t \end{aligned}$$

因为随机变量的数学期望值是随机变量的最佳估计值，所以可取 $s_t^{(1)}$ 、$s_t^{(2)}$ 代替 $E(s_t^{(1)})$、$E(s_t^{(2)})$，从而有

$$\begin{cases} s_t^{(1)} = a_t - \dfrac{1-\alpha}{\alpha}b_t \\ s_t^{(2)} = a_t - \dfrac{2(1-\alpha)}{\alpha}b_t \end{cases}$$

由此解得

$$\begin{cases} a_t = 2s_t^{(1)} - s_t^{(2)} \\ b_t = \dfrac{\alpha}{1-\alpha}(s_t^{(1)} - s_t^{(2)}) \end{cases} \tag{4-25}$$

【例 4-5】 某港口城市平均每人商业销售额的数据如表 4-5 中第(2)栏所示，用二次指数平滑法(α 取 0.8)计算历年的理论预测值。

表 4-5 二次指数平滑法计算表 单位：元

年份	t	商业销售额 y_t	$s_t^{(1)}$	$s_t^{(2)}$	$s_t^{(1)}-s_t^{(2)}$	a_t	b_t	y_{t+1} $T=1$
(1)		(2)	(3)	(4)	(5)=(3)−(4)	(6)=(3)+(5)	(7)=(4)×(5)	(8)=(6)+(7)
2002	1	243.29	243.2	243.2	0	243.29	0	
2003	2	277.82	270.91	265.39	5.52	276.43	22.08	243.29
2004	3	320.39	310.49	301.47	9.02	319.51	36.08	298.51
2005	4	389.09	373.37	358.99	14.38	387.75	57.52	355.59
2006	5	444.84	430.55	416.24	14.31	444.86	57.244	45.27
2007	6	496.23	483.09	469.72	13.37	496.46	53.48	502.1
2008	7							549.94

【解】 平滑法的步骤类似于二次移动平均法的步骤，需注意的是，在二次指数平滑法中 $s_t^{(1)}$ 和 $s_t^{(2)}$ 不是预测值，而是为计算最终预测值服务的平均值，要注意其公式和一次指数平滑法公式的差别，并且要注意 $s_t^{(1)}$ 和 $s_t^{(2)}$ 的值在计算表中的位置。例如：

$$s_6^{(1)} = ay_6 + (1-a)s_5^{(1)} = 0.8 \times 496.23 + 0.2 \times 430.55 = 483.09(\text{元})$$

所得数值相应地放在 $t=6$ 的位置上

$$a_6 = s_6^{(1)} + (s_6^{(1)} - s_6^{(2)}) = 483.09 + 13.37 = 496.48$$

$$b_6 = \frac{\alpha}{1-\alpha}(s_6^{(1)} - s_6^{(2)}) = 4 \times 13.37 = 53.48$$

其余类似。

2008 年的预测值为

$$\hat{y}_{6+1} = a_6 + b_6 \times 1 = 496.46 + 53.48 = 549.94(元)$$

若要预测 2010 年的值，则

$$\hat{y}_{6+3} = a_6 + b_6 \times 3 = 496.46 + 53.48 \times 3 = 656.9(元)$$

4.3.3 三次指数平滑

当数据模型有二次、三次或高次幂时，即具有非线性趋势时，需要用高次平滑形式。当时间序列的观察值经二次指数平滑处理后的时间序列仍有曲率，则原有时间序列进行三次指数平滑，用二次曲线来描述。三次指数是在二次指数平滑的基础上，再进行一次指数平滑，其计算公式为

$$\begin{aligned} s_t^{(1)} &= \alpha y_t + (1-\alpha)s_{t-1}^{(1)} \\ s_t^{(2)} &= \alpha s_t^{(1)} + (1-\alpha)s_{t-1}^{(2)} \\ s_t^{(3)} &= \alpha s_t^{(2)} + (1-\alpha)s_{t-1}^{(3)} \end{aligned} \tag{4-26}$$

式中，$s_t^{(3)}$ 为三次指数平滑值。

三次指数平滑法的预测模型为

$$\hat{y}_{t+T} = a_t + b_t T + c_t T^2 \tag{4-27}$$

式中，

$$\begin{cases} a_t = 3s_t^{(1)} - 3s_t^{(2)} + s_t^{(3)} \\ b_t = \dfrac{\alpha}{2(1-\alpha)^2}[(6-5\alpha)s_t^{(1)} - 2(5-4\alpha)s_t^{(2)} + (4-3\alpha)s_t^{(3)}] \\ c_t = \dfrac{\alpha^2}{2(1-\alpha)^2}[s_t^{(1)} - 2s_t^{(2)} + s_t^{(3)}] \end{cases} \tag{4-28}$$

三次指数平滑预测不仅考虑了时间序列线性增长的因素，也考虑了二次曲线的增长因素，因此对预测对象是二次曲线趋势的时间序列较好的预测方法。与二次指数平滑类似也可以做期数以后的预测，但是也只能适用于短、近期预测，不能用于中、长期预测。

【例 4-6】 某港务集团公司 2001—2011 年收入总额如表 4-6 所示，当 $\alpha =$

0.3 时，试预测 2012、2014 年销售收入各为多少万元？

表 4-6 为港务集团公司 2001—2011 年收入总额及一次、二次、三次平滑值计算表。

表 4-6 指数平滑值计算表 单位：万元

年 份	t	收入总额 y_t	一次平滑值 $s_t^{(1)}$	二次平滑值 $s_t^{(2)}$	三次平滑值 $s_t^{(3)}$	$\hat{y}_{t+1}$
2001	1	2 004	2 137.0	2 176.9	2 188.9	2 194.0
2002	2	2 006	2 097.8	2 178.2	2 178.2	2 023.1
2003	3	2 572	2 240.0	2 178.5	2 178.5	1 956.6
2004	4	3 461	2 606.3	2 217.1	2 217.1	2 449.5
2005	5	5 177	3 377.5	2 340.5	2 340.5	3 459.6
2006	6	5 592	4 041.9	2 554.1	2 554.1	5 388.9
2007	7	8 065	5 248.8	3 711.4	2 901.3	6 457.9
2008	8	13 111	6 707.5	4 880.2	3 495	8 929.4
2009	9	14 858	9 782.6	6 350.9	4 351.7	14 242.7
2010	10	16 267	11 727.9	7 964.0	5 435.4	17 608.6
2011	11	23 226	15 177.4	10 128.0	6 843.2	19 626.0
2012						25 991.8

【解】 (1) 绘制散点图。由图 4-3 可知，其年销售收入是二次曲线上升，故可用三次指数平滑法进行预测。

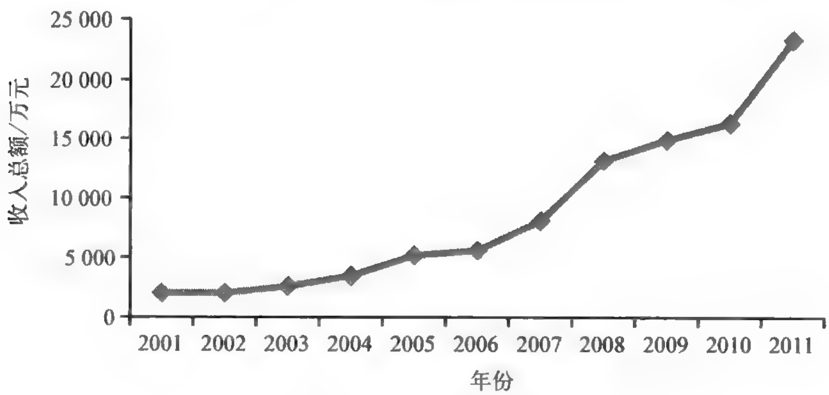


图 4-3 2001—2011 年收入总额

(2) 选择指数平滑系数。本题 α 取 0.3。

(3) 确定初始值。

$$s_0^{(1)} = s_0^{(2)} = s_0^{(3)} = \frac{y_1 + y_2 + y_3}{3} = \frac{2\,004 + 2\,006 + 2\,572}{3} = 2\,194$$

(4) 计算一次、二次、三次指数平滑值。

$$s_1^{(1)} = ay_1 + (1-a)s_0^{(1)} = 0.3 \times 2\,004 + (1-0.3) \times 2\,194 = 2\,137$$

……

$$s_{11}^{(1)} = ay_{11} + (1-a)s_{10}^{(1)} = 0.3 \times 23\,226 + (1-0.3) \times 11\,727.9 = 15\,177.4$$

$$s_1^{(2)} = as_{11}^{(1)} + (1-a)s_0^{(2)} = 0.3 \times 2\,137 + (1-0.3) \times 2\,194 = 2\,176.9$$

……

$$s_{11}^{(2)} = as_{11}^{(1)} + (1-a)s_{10}^{(2)} = 0.3 \times 15\,177.4 + (1-0.3) \times 764.0 = 10\,128$$

$$s_1^{(3)} = as_1^{(2)} + (1-a)s_0^{(3)} = 0.3 \times 2\,176.9 + (1-0.3) \times 2\,194 = 2\,188.9$$

……

$$s_{11}^{(3)} = as_{11}^{(2)} + (1-a)s_{10}^{(3)} = 0.3 \times 10\,128 + (1-0.3) \times 5\,435.4 = 6\,843.2$$

(5) 计算待定参数,建立预测模型。

$$a_t = 3s_1^{(1)} - 3s_t^{(2)} + 3s_t^{(3)} = 3 \times 15\,177.4 - 3 \times 10\,128 + 6\,843.2 = 21\,991.2$$

$$\begin{aligned} b_t &= \frac{\alpha}{2 \times (1-\alpha)^2}[(6-5\alpha)s_t^{(1)} - 2(5-4\alpha)s_t^{(2)} + (4-3\alpha)s_t^{(3)}] \\ &= \frac{0.3}{2 \times (1-0.3)^2}[(6-5 \times 0.3) \times 15\,177.4 - 2 \times (5-4 \times 0.3) \times \\ &\quad 10\,128 + (4-3 \times 0.3) \times 6\,843.2] \\ &= 3\,838.6 \end{aligned}$$

$$\begin{aligned} c_t &= \frac{\alpha^2}{2\,(1-\alpha)^2}[s_t^{(1)} - 2s_t^{(2)} + s_t^{(3)}] \\ &= \frac{0.3^2}{2 \times (1-0.3)^2} \times [15\,177.4 - 2 \times 10\,128 + 6\,843.2] \\ &= 162 \end{aligned}$$

故三次指数平滑法预测模型为

$$\hat{y}_{t+T} = 21\,991.2 + 3\,838.6T + 162T^2$$

2012 年、2014 年的销售收入预测值分别为

$$\hat{y}_{11+1} = 21\,991.2 + 3\,838.6 \times 1 + 162 \times 1^2 = 25\,991.8(万元)$$

$$\hat{y}_{11+3} = 21\,991.2 + 3\,838.6 \times 3 + 162 \times 3^2 = 34\,965.0(万元)$$

4.4 差分-指数平滑法

时间序列按直线趋势变动时，用一次指数平滑法预测时，由于数据不满足模型要求，会出现滞后偏差，所以，考虑到先对数据进行变换，使之能适合于一次指数平滑模型，然后在对输出结果做技术上的返回处理，使之恢复为原变量的形态。差分方法是改变数据变动趋势的简易方法。下面讨论差分的方法来改进指数平滑法。

4.4.1 一阶差分-指数平滑模型

当时间序列具有直线趋势变动时，可运用一价差分-指数平滑模型来预测。其公式如下：

$$\nabla y_t = y_t - y_{t-1} \tag{4-29}$$

$$\nabla \hat{y}_{t+1} = \alpha \nabla y_t + (1-\alpha) \nabla \hat{y}_t \tag{4-30}$$

$$\hat{y}_{t+1} = \nabla \hat{y}_{t+1} + y_t \tag{4-31}$$

式中，∇是差分记号。

式(4-29)表示对呈现直线增加的序列做一价差分，构成一个平稳的新序列；式(4-30)表示对这个新序列运用指数平滑法；式(4-31)表示把经过一价差分后的新序列的指数平滑预测值与变量当前的实际值叠加，作为变量下一期的预测值。对于这个公式的数学意义可做如下的解释：

$$y_{t+1} = y_{t+1} - y_t + y_t = \nabla y_{t+1} + y_t \tag{4-32}$$

但是，在 t 为当前期时，y_{t+1} 实际上还没有，因此不能按照式(4-29)来计算 ∇y_{t+1}，而只能进行估计。我们采用按式(4-30)计算的预测值 $\nabla \hat{y}_{t+1}$ 去估计式(4-32)中的 ∇y_{t+1}，从而式(4-32)等号左项的 y_{t+1} 也要改为预测值 $\hat{y}_{t+1}$，亦即成为式(4-31)。

在前面我们已分析过，指数平滑值实质上是一种加权平均数。因此把序列中逐期增量的加权平均数(指数平滑值)加上当前期的实际数进行预测，它比一次指数平滑法只用变量以往取值的加权平均数作为下一期的预测更为合理。从而使预测值始终围绕实际值上下波动，根本上克服了在有直线增长趋势的情况下，用一次指数平滑法所取得出的结果始终落后于实际值的弊端。

【例 4-7】 某港口集团下属码头公司 2002—2011 年机械燃油消耗量资料如表 4-7 所示，试预测 2012 年的燃料消耗量。

表 4-7 某企业机械燃料消耗的差分指数平滑法计算表 ($\alpha = 0.4$) 单位：百吨

年份	t	燃料消耗量 y_t	差分 $\nabla y_t = y_t - y_{t-1}$	差分指数平滑值 $\nabla \hat{y}_{t+1} = \alpha \nabla y_t + (1-\alpha) \nabla \hat{y}_t$	预测值 $\hat{y}_{t+1} = \nabla \hat{y}_{t+1} + y_t$
2002	1	24			
2003	2	26	2		
2004	3	27	1	2	28
2005	4	30	3	1.6	28.6
2006	5	32	2	2.16	32.16
2007	6	33	1	2.10	34.10
2008	7	36	3	1.66	34.66
2009	8	40	4	2.20	38.20
2010	9	41	1	2.92	42.92
2011	10	44	3	2.15	43.15
2012	11	—		2.49	46.49

【解】 由表 4-7 可看出，燃料消耗量，除个别年份外，逐期增长量大体在 200 吨左右，即呈直线增长，因此可用一价差分-指数平滑模型来预测。我们取 $\alpha = 0.4$，预测 2012 年燃料消耗量为

$$\hat{y}_{2007} = 2.49 + 44 = 46.49(\text{百吨})$$

4.4.2 二阶差分-指数平滑模型

当时间序列呈现二次曲线增长时，可用二次差分-指数平滑模型来预测，其公式如式(4-33)～式(4-36)所示：

$$\nabla y_t = y_t - y_{t-1} \tag{4-33}$$

$$\nabla^2 y_t = \nabla y_t - \nabla y_{t-1} \tag{4-34}$$

$$\nabla^2 \hat{y}_{t+1} = \alpha \nabla^2 y_t + (1-\alpha) \nabla^2 \hat{y}_t \tag{4-35}$$

$$\hat{y}_{t+1} = \nabla^2 \hat{y}_{t+1} + \nabla y_t + y_t \tag{4-36}$$

式中，∇^2 表示二价差分，与一价差分-指数平滑模型类似。

因为

$$\begin{aligned} y_{t+1} &= y_{t+1} - y_t + y_t \\ &= \nabla y_{t+1} + y_t \\ &= (\nabla y_{t+1} - \nabla y_t) + \nabla y_t + y_t \\ &= \nabla^2 y_{t+1} + \nabla y_t + y_t \end{aligned}$$

同样，用 $\nabla^2 \hat{y}_{t+1}$ 代替 $\nabla^2 y_{t+1}$ 得到式(4-36)：

$$\hat{y}_{t+1} = \nabla^2 \hat{y}_{t+1} + \nabla y_t + y_t$$

差分方法和指数平滑法的联合运用，除了能克服一次指数平滑法的滞后偏差外，对初始值的问题也有显著的改进。因为数据经差分平稳化处理后，所产生的新序列基本上是平稳的。这时，初始值取新序列的第一期数据对于未来预测值不会有多大影响。其次，它开拓了指数平滑法的适应范围，使一些原来需要运用配合趋势线方法处理的情况可用这种组合模型来取代。但是，对于指数平滑法存在的加权系数 α 的选择问题，以及只能逐期预测问题，差分-指数平滑模型也没能做出改进。

4.5 自适应过滤法

4.5.1 自适应过滤法的基本过程

自适应过滤法与移动平均法、指数平滑法一样，也是以时间序列的历史观察值进行某种加权平均来预测的，它要寻找一组“最佳”的权数，其办法是先用一组给定的权数来计算一个预测值，然后计算预测误差，再根据预测误差调整权数以减小误差。这样反复进行，直至找出一组“最佳”权数，使误差减少到最低限度。由于这种调整权数的过程与通信工程中过滤传输噪声的过程极为接近，故称为自适应过滤法。

自适应过滤法的基本预测公式为

$$\hat{y}_{t+1} = \omega_1 y_t + \omega_2 y_{t-1} + \cdots + \omega_n y_{t-n+1} = \sum_{i=1}^{n} \omega_i y_{t-i+1} \tag{4-37}$$

式中，$\hat{y}_{t+1}$ 为第 $t+1$ 期的预测值；ω_i 为第 $t-i+1$ 期的观察值权数；y_{t-i+1} 为第 $t-i+1$ 期的观察值；n 为权数的个数。

其调整权数的公式为

$$\omega_i' = \omega_i + 2ke_{t+1} y_{t-i+1} \tag{4-38}$$

式中，$i=1, 2, \cdots, N$；$t=N, N+1, \cdots, n$，n 为序列数据的个数；ω_i 为调整前的第 i 个权数；ω_i' 为调整后的第 i 个权数；k 为学习常数；e_{t+1} 为第 $t+1$ 期的预测误差。

式(4－38)表明：调整后的一组权数应等于旧的一组权数加上误差调整项，这个调整项包括预测误差、原观察值和学习常数等三个因素。学习常数 k 的大小决定权数调整的速度。

下面举一个简单的例子来说明此法的全过程。设有一个时间序列包括 10 个观察值。如表 4－8 所示。

表 4－8 某时间序列表

时间 t	1	2	3	4	5	6	7	8	9	10
观察值 y_t	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0

试用自适应过滤法，以两个权数来求第 11 期的预测值。

本例中：$N=2$，取初始权数 $\omega_1=0.5$，$\omega_2=0.5$，并设 $k=0.9$，t 的取值由 $N=2$ 开始。

当 $t=2$ 时：

(1) 按预测公式(4－37)求第 $t+1=3$ 期的预测值：

$$\begin{aligned}\hat{y}_{t+1}&=\hat{y}_3=\omega_1 y_2+\omega_2 y_1\\&=0.5\times0.2+0.5\times0.1=0.15\end{aligned}$$

(2) 计算预测误差：

$$\begin{aligned}e_{t+1}&=e_3=y_3-\hat{y}_3\\&=0.3-0.15=0.15\end{aligned}$$

(3) 根据式(4－38) $\omega_i'=\omega_i+2ke_{t+1}y_{t-i+1}$ 调整权数：

$$\begin{aligned}\omega_1'&=\omega_i+2ke_3y_2\\&=0.5+2\times0.9\times0.15\times0.2=0.554\\\omega_2'&=\omega_2+2ke_3y_1\\&=0.5+2\times0.9\times0.15\times0.1=0.527\end{aligned}$$

这(1)～(3)结束，即完成了一次权数调整，然后 t 进 1 再重复以前步骤。

当 $t=3$ 时：

(1) 利用所得到的权数，计算第 $t+1=4$ 期的预测值。方法是，舍去最前面的一个观察值 y_1，增加一个新的观察值 y_3 即

$$\begin{aligned}\hat{y}_{t+1} &= \hat{y}_4 = \omega_1 y_3 + \omega_2 y_2 \\ &= 0.554 \times 0.3 + 0.527 \times 0.2 = 0.2716\end{aligned}$$

(2) 计算预测误差：

$$\begin{aligned}e_{t+1} &= e_4 = y - \hat{y}_4 \\ &= 0.4 - 0.2716 = 0.1284 \approx 0.13\end{aligned}$$

(3) 调整权数：

$$\omega'_1 = 0.554 + 2 \times 0.9 \times 0.13 \times 0.3 = 0.624$$
$$\omega'_2 = 0.527 + 2 \times 0.9 \times 0.13 \times 0.2 = 0.564$$

这样进行到 $t = 10$ 时，$\hat{y}_{t+1} = \hat{y}_{11} = \omega'_1 y_{10} + \omega'_2 y_9$，但由于没有 $t = 11$ 时的观察值 y_{11}，因此 $e_{t+1} = e_{11} = y_{11} - \hat{y}_{11}$ 无法计算。这时，第一轮的调整就此结束。把现有的新权数作为初始权数，重新开始 $t = 2$ 时的过程。这样反复进行下去，到预测误差(指一轮预测的总误差)没有多大改进时，就认为获得了一组“最佳”权数，能实际用来预测第 11 期的数值。本例在调整过程中，可使误差降到零，而权数达到稳定不变，最后得到的“最佳”权数为 $\omega'_1 = 2.0$，$\omega'_2 = -1.0$，用“最佳”权数预测第 11 期的取值：

$$\hat{y}_{11} = \omega'_1 y_{10} + \omega'_2 y_9 = 2 \times 0.1 + (-1) \times 0.9 = 1.1$$

在实际应用中要达到这样的结果，调整计算的工作量可能很大，必须借助于计算机才能容易实现。

4.5.2 N, k 值和初始权数的确定

在开始调整权数时，首先要确定权数个数 N 和学习常数 k。一般来说，当时间序列的观察值呈现季节变动时，N 应取季节性长度值。如序列以一年为周期进行季节变动时，若数据是月度的，则取 $N = 12$；若数据是季度的，则取 $N = 4$。如果时间序列无明显的周期变动，则可用自相关系法来确定，即用最高自相关系数的滞后时期数定作 N。

k 的取值一般可定为 $1/N$；也可以用不同的 k 值进行试算，以确定一个能使 MSE 最小的 k 值。

初始权数的确定也很重要，如无其他依据，也可用 $1/N$ 作初始权系数用，即 $\omega_i = 1/N(i = 1, 2, 3, \cdots, N)$。

自适应过滤法有两个明显的优点：一是技术比较简单，可根据预测意图来选择权数的个数和学习常数，以控制预测。也可以由计算机自动选定。二是它使用

了全部历史数据来寻求最佳权系数，并随数据轨迹的变化而不断更新权数，从而不断改进预测。

思考与练习

1. 移动平均预测法有哪些不足？
2. 什么是二次指数平滑法？它与二次移动平均法相比有何优势？
3. 如何确定一次指数平滑法的初始值？如何选择平滑系数 α？
4. 何时运用三次指数平滑法？其预测步骤有哪些？
5. 某市 2012 年 1—11 月份某商品的销售额如题表 4－1 所示，n 分别取 3、5（个月），用一次移动平均法预测 12 月份的销售额，并比较其优劣。（单位：万元）

题表 4－1

月份	1	2	3	4	5	6	7	8	9	10	11	12
销售额	400	270	380	396	620	350	310	260	440	540	470	

6. 某公司 2012 年 6—11 月出口货物数量如题表 4－2 所示，一次给定权数为 0.5、1.0、1.5、2.0、2.5、3.0，试用加权平均法预测 12 月份出口货物量。（单位：吨）

题表 4－2

月份	6	7	8	9	10	11	12
出口量	19	18	19	21	20	22	

7. 某码头 2000—2011 年吞吐量如题表 4－3 所示，用二次移动平均法预测（$n=4$）2012 年该码头吞吐量。（单位：万吨）

题表 4－3

年份	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
吞吐量	192	224	188	198	206	203	238	228	231	221	259	273

8. 某企业 2004—2010 年出口额如题表 4－4 所示，用二次指数平滑法（$\alpha=0.8$）预测该企业 2011 年和 2012 年的出口额。（单位：万元）

题表 4－4

年份	2004	2005	2006	2007	2008	2009	2010
出口额	300	324	347	372	396	420	446

9. 某港口 1990—2012 年的吞吐量如题表 4-5 所示，试用三次指数平滑法预测 2013、2014 和 2015 年的吞吐量各位多少？（单位：万吨）

题表 4-5

年份	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
吞吐量	697	608	604	605	638	670	732	770	738	802	858	930	1 023
年份	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012			
吞吐量	1 106	1 164	1 271	1 340	1 434	1 558	1 800	2 140	2 350	2 570			

5 趋势曲线模型预测法

移动平均法和指数平滑法主要适用于短期预测，而长期趋势预测法则是进行中长期预测的主要方法。它是根据时间序列的发展趋势，配合合适的曲线模型，外推预测未来的趋势值，对未来状况做出预测的方法。这种方法是以预测的连续性原理为基础的，通过对有关预测对象的历史数据的抽象和现有状况的分析找出其发展的规律性，对预测对象未来状况做出预测。长期趋势预测法的各种预测模型都是遵循此原理建立并应用于实际中的。

5.1 直线模型预测法

直线模型预测法是根据预测对象具有线性变动趋势的历史数据，拟合成一条直线，通过建立直线模型进行预测的方法。它是长期趋势预测法的基本方法，也是预测实践中最常用的方法。

直线预测模型为

$$\hat{y} = a + bt \tag{5-1}$$

式中，t 为时间，代表年次、月次等；$\hat{y}$ 多为预测值；a、b 为参数，a 代表 $t=0$ 时的预测值，b 代表逐期增长量。

直线预测模型的特点，是一阶差分为一常数：

$$\nabla \hat{y}_t = \hat{y}_t - \hat{y}_{t-1} = b$$

因此，当时间序列 $\{y\}$ 的一阶差分 ∇y_t 近似为一常数，其散点图呈直线趋势时，可配合直线预测模型来预测。

直线预测模型的参数，可用最小平方法、折扣最小平方法等来估计。

5.1.1 最小平方法

最小平方法，就是使各期的实际值与预测值之间的误差平方和 $Q = \sum (y_t - \hat{y}_t)^2$，即 $Q = \sum (y_t - a - bt)^2$ 达到最小来估计 a 和 b 的方法。

由极值原理，要使 Q 最小，则必须满足：

$$\frac{\partial Q}{\partial a}=-2\sum(y_t-a-bt)=0$$

$$\frac{\partial Q}{\partial b}=-2\sum(y_t-a-bt)\cdot t=0$$

整理可得标准方程组：

$$\begin{cases}\sum y_t=na+b\sum t\\ \sum ty_t=a\sum t+b\sum t^2\end{cases}\tag{5-2}$$

式中，n 为时间序列的项数。

为了简化计算，可选取时间序列 $\{y_t\}$ 的中点为时间原点，使 $\sum t=0$。当序列数为奇数项时，t 分别为…，-2，-1，0，1，2，…；当序列数为偶数项时，t 分别为…，-5，-3，-1，1，3，5，…。则方程组(5-2)可简化为

$$\begin{cases}\sum y_t=na\\ \sum ty_t=b\sum t^2\end{cases}\tag{5-3}$$

由此可得

$$\begin{cases}\hat{a}=\dfrac{\sum y_t}{n}\\ \hat{b}=\dfrac{\sum ty_t}{\sum t^2}\end{cases}\tag{5-4}$$

【例 5-1】 2007—2012 年某码头公司大学本科以上学历的从业人员在从业人员总数中所占的比例如表 5-1 所示，试用最小平方法求参数并预测 2013 年大学本科以上学历从业人员所占的比例。

表 5-1 大学本科以上学历从业人员所占的比例及其一阶差分 单位：%

年 份	2007	2008	2009	2010	2011	2012
比 例	18.9	19.8	21.2	23.0	24.8	26.0
一阶差分 ∇y_i		0.9	1.4	1.8	1.8	1.2

【解】

(1) 选择预测模型。

计算序列的一阶差分，列于表 5-1 中，从计算结果可以看出，一阶差分大体接近。因此，可配合直线预测模型 $\hat{y}_t = a + bt$ 来预测。

(2) 列表计算求待定系数所需的数据资料。

以 $\sum t = 0$ 对 t 进行编号，$t = -5, -3, -1, 1, 3, 5$。

根据求参数需要，列最小平方法求参数计算表，如表 5-2 所示。

表 5-2 最小平方法求参数计算表

年 份	t	y_t	ty_t	t^2
2007	−5	18.9	−94.5	25
2008	−3	19.8	−59.4	9
2009	−1	21.2	−21.2	1
2010	1	23.0	23.0	1
2011	3	24.8	74.4	9
2012	5	26.0	130.0	25
$\sum$	0	133.7	52.3	70

因为 $\sum t = 0$，所以 $a = \dfrac{\sum y_t}{n} = \dfrac{133.7}{6} \approx 22.3$，$b = \dfrac{\sum t y_t}{t^2} = \dfrac{52.3}{70} \approx 0.7$

可得预测模型为

$$\hat{y}_t = a + bt = 22.3 + 0.7t$$

将各年次的 t 值代入预测模型，可得各年的追溯预测值 $\hat{y}_t$。

(3) 预测。

令 $t = 7$，可得 2013 年的预测值为

$$\hat{y}_t = 22.3 + 0.7 \times 7 = 27.2\%$$

5.1.2 折扣最小平方法

最小平方法是估计线性模型参数的常用方法，但是它有个缺陷，就是把近期误差与远期误差的重要性等同看待。实际上，近期误差比远期误差重要得多。为了克服这个缺陷，在预测中，常采用折扣最小平方法，进行合理的加权，对近期误差比

对远期误差给以较大的权数。

折扣最小平方法就是对误差平方进行指数折扣加权后，使其总和达到最小的方法。其数学表达式为

$$Q=\sum_{t=1}^{n}\alpha^{n-t}(y_t-\hat{y}_t)^2=\min \tag{5-5}$$

式中，α 为折扣系数，$0<\alpha<1$。

由式(5－5)可以看出，最近期的误差平方 $(y_n-\hat{y}_n)^2$ 的权数为 α^0，最远期的误差平方 $(y_n-\hat{y}_n)^2$ 的权数为 α^{n-1}。第 t 期的误差平方的权数为 α^{n-t}。由于 $\alpha^0=1,\cdots,\alpha^{n-t},\cdots,\alpha^{n-1}$ 是越来越小的权数，这说明对最近期的误差平方不打折扣，而对远期的误差平方越远打的折扣越大，所以称为折扣最小平方法。折扣的程度视 α 取值大小而异，α 值越接近 0，折扣加权作用越大；α 值越接近 1，折扣加权作用越小。如 $\alpha=1$，则指数折扣加权就失去了作用，作为不打折扣的最小平方法了。

下面我们用折扣最小平方法来估计直线预测模型 $\hat{y}_t=a+bt$ 的参数 a、b，使

$$Q=\sum_{t=1}^{n}\alpha^{n-t}(y_t-a-bt)^2=\min \tag{5-6}$$

对式(5－6)求偏导数，便求得参数 a、b 估计值的标准方程组为

$$\begin{cases}\sum\limits_{t=1}^{n}\alpha^{n-t}y_t=a\sum\limits_{t=1}^{n}\alpha^{n-t}+b\sum\limits_{t=1}^{n}a^{n-t}t\\ \sum\limits_{t=1}^{n}\alpha^{n-t}ty_t=a\sum\limits_{t=1}^{n}\alpha^{n-t}t+b\sum\limits_{t=1}^{n}a^{n-t}t^2\end{cases} \tag{5-7}$$

【例 5－2】 某第三方物流公司 2001～2011 年利润额数据资料如表 5－3 所示，求当 $\alpha=0.8$ 时，用折扣最小平方法预测 2012 年的利润额是多少万元？

【解】 (1) 列表计算有关数据，计算结果如表 5－3 所示。

表 5－3 折扣最小平方法计算表 单位：万元

年份	t	利润额 y_t	$n-t$	α^{n-t}	$\alpha^{n-t}y_t$	$\alpha^{n-t}ty_t$	$\alpha^{n-t}t$	$\alpha^{n-t}t^2$	$\hat{y}_t$	$(y_t-\hat{y}_t)^2$	$\alpha^{n-t}\cdot(y_t-\hat{y}_t)$
2001	1	200	10	0.107 4	21.48	21.48	0.107 4	0.107 4	185.34	214.92	23.08
2002	2	300	9	0.134 2	40.27	80.52	0.268 4	0.536 8	269.00	961.00	128.97
2003	3	350	8	0.167 8	58.72	176.19	0.503 4	1.510 2	352.66	7.08	1.19
2004	4	400	7	0.209 7	83.87	335.52	0.838 8	3.355 2	436.32	1 319.14	276.62
2005	5	500	6	0.262 1	131.07	655.25	1.310 5	6.552 5	519.98	399.20	104.63

续 表

年份	t	利润额 y_t	$n-t$	α^{n-t}	$\alpha^{n-t}y_t$	$\alpha^{n-t}ty_t$	$\alpha^{n-t}t$	$\alpha^{n-t}t^2$	$\hat{y}_t$	$(y_t-\hat{y}_t)^2$	$\alpha^{n-t}\cdot(y_t-\hat{y}_t)$
2006	6	630	5	0.327 7	206.44	1 238.71	1.966 2	11.792 2	603.64	694.85	227.70
2007	7	700	4	0.409 6	286.72	2 007.04	2.867 2	20.070 4	687.30	161.29	66.06
2008	8	750	3	0.512 0	384.00	3 072.00	4.096 0	32.768 0	770.96	439.32	224.93
2009	9	850	2	0.640 0	544.00	4 896.00	5.760 0	51.840 0	854.62	21.34	13.66
2010	10	950	1	0.800 0	760.00	7 600.00	8.000 0	80.000 0	938.28	137.36	109.88
2011	11	1 020	0	1.000 0	1 020.00	11 220.00	11.000 0	121.000 0	1 021.94	3.76	3.76
$\sum$				4.570 5	3 536.57	31 302.71	36.717 9	329.537 7			1 180.48

将计算结果代入式(5-7)得

$$3\,536.57 = 4.570\,5a + 36.717\,9b$$
$$31\,302.71 = 36.717\,9a + 329.537\,7b$$

解此方程组得 $\hat{a} = 101.68,\ \hat{b} = 83.66$。

所求直线预测模型为

$$\hat{y}_t = 101.68 + 83.66t$$

(2) 求预测值。

将各年的 t 值代入预测模型,可得各年的追溯预测值 $\hat{y}$,如表 5-3 所示。追溯预测值是比较准确的。

当 $t = 12$ 时,可得 2012 年的预测值:

$$\hat{y}_{12} = 101.68 + 83.66 \times 12 = 1\,105.6\ (\text{万元})$$

(3) 计算估计标准误差。

$$S_y = \sqrt{\frac{\sum \alpha^{n-t}(y_t - \hat{y}_t)^2}{n-m}} \tag{5-8}$$

式中,n 为资料总项数;m 为模型参数的个数。将表 5-3 中的结果代入,得

$$S_y = \sqrt{\frac{\sum \alpha^{n-t}(y_t - \hat{y}_t)^2}{n-m}} = \sqrt{\frac{1\,180.48}{11-2}} = 11.45$$

可以看出,由于折扣最小平方法按重近轻远的赋值原则,使其预测值更接近实

际观察值，而且，随 α 取值越小越突出。这是因为 α 取值越小，对近期的权数就显得越大，因而预测值就越接近实际观察值，这就是折扣最小平方法的作用。但是，如何选取合适的折扣系数 α，却是一件比较麻烦的事。一般可利用电子计算机，在 0～1 的范围内，选取若干 α 值进行试算，使 $Q=\sum_{t=1}^{n}\alpha^{n-t}(y_t-\hat{y}_t)^2$ 达到最小的那个 α 为最好。

5.2 多项式曲线模型预测法

在很多情况下，自变量与因变量的关系由于受众多因素的影响，其变动趋势并非总是一条简单的直线方程，往往会呈现不同形态的曲线变动趋势，可用多项式曲线模型来表示。

多项式曲线预测模型的一般形式为

$$\hat{y}_t=a+bt+ct^2+dt^3+et^4+\cdots \tag{5-9}$$

直线预测模型 $\hat{y}=a+bt$ 是它的特殊形式。这一节主要介绍常用的二次和三次抛物线的预测模型。

二次抛物线预测模型为

$$\hat{y}_t=a+bt+ct^2 \tag{5-10}$$

式中，a、b、c 为参数；t 为时间。

二次抛物线预测模型的特点是二阶差分为一常数：

$$\nabla^2\hat{y}_t=\nabla\hat{y}_t-\nabla\hat{y}_{t-1}=2c$$

其图形是一个弯曲向上凸或向下凹的曲线。因此，当时间序列 $\{y_t\}$ 的二阶差分 $\nabla^2 y_t$ 近似为一个常数，其散点图呈现一个弯曲向上凸或向下凹的曲线发展趋势时，可配合二次抛物线预测模型来预测。

三次抛物线预测模型为

$$\hat{y}_t=a+bt+ct^2+dt^3 \tag{5-11}$$

式中，a、b、c、d 为参数；t 为时间。

三次抛物线预测模型的特点是三阶差分为一常数：

$$\nabla^3\hat{y}_t=\nabla^2\hat{y}_t-\nabla^2\hat{y}_{t-1}=6d$$

其图形是一条有两个弯曲的曲线。因此，当时间序列 $\{y_t\}$ 的三阶差分 $\nabla^3\hat{y}_t$ 近似为

一常数，其散点图呈现两个弯曲曲线的发展趋势时，可配合三次抛物线预测模型来预测。

二次和三次抛物线预测模型的参数，可用最小平方法、三点法等来估计。

5.2.1 最小平方法

以三次抛物线预测模型的参数估计为例。

$$\hat{y}_t = a + bx + cx^2 + dx^3$$

根据最小平方法的原理，令 $\dfrac{\partial Q}{\partial a} = \dfrac{\partial Q}{\partial b} = \dfrac{\partial Q}{\partial c} = \dfrac{\partial Q}{\partial d} = 0$，可得标准方程组：

$$\begin{cases} \sum y_t = na + b\sum t + c\sum t^2 + d\sum t^3 \\ \sum ty_t = a\sum t + b\sum t^2 + c\sum t^3 + d\sum t^4 \\ \sum t^2 y_t = a\sum t^2 + b\sum t^3 + c\sum t^4 + d\sum t^5 \\ \sum t^3 y_t = a\sum t^3 + b\sum t^4 + c\sum t^5 + d\sum t^6 \end{cases} \tag{5-12}$$

为了简化计算，可选取时间序列 $\{y_t\}$ 的中点为时间原点，使 $\sum t = 0$，$\sum t^3 = 0$，$\sum t^5 = 0$，从而使上列方程组简化为

$$\begin{cases} \sum y_t = na + c\sum t^2 \\ \sum ty_t = b\sum t^2 + d\sum t^4 \\ \sum t^2 y_t = a\sum t^2 + c\sum t^4 \\ \sum t^3 y_t = b\sum t^4 + d\sum t^6 \end{cases} \tag{5-13}$$

由此可解得 a、b、c、d 的估计值。

【例 5-3】 某公司 1998—2010 年商品出口额如表 5-4 所示，试预测 2011 年出口额为多少万元？

表 5-4 某公司商品出口额及三次抛物线预测模型最小平方法计算表 单位：万元

年份	t	出口额 y_t	t^2	t^3	t^4	t^6	ty_t	t^2y_t	t^3y_t	$\hat{y}_t$	$(y_t-\hat{y}_t)^2$
1998	−6	150	36	−216	1 296	46 656	−900	5 400	−32 400	133.81	262.18
1999	−5	165	25	−125	625	15 625	−825	4 125	−20 625	175.38	107.24

续 表

年份	t	出口额 y_t	t^2	t^3	t^4	t^6	ty_t	t^2y_t	t^3y_t	$\hat{y}_t$	$(y_t-\hat{y}_t)^2$
2000	−4	180	16	−64	256	4 096	−720	2 880	−11 520	202.44	503.55
2001	−3	210	9	−27	81	729	−630	1 890	−5 670	217.43	55.8
2002	−2	225	4	−8	16	64	−450	900	−1 800	222.95	4.22
2003	−1	240	1	−1	1	1	−240	240	−240	221.37	347.08
2004	0	240	0	0	0	0	0	0	0	215.25	612.79
2005	1	210	1	1	1	1	210	210	210	207.05	8.73
2006	2	195	4	8	16	64	390	780	1 560	199.28	18.28
2007	3	165	9	27	81	729	495	1 484	4 455	194.42	865.24
2008	4	195	16	64	256	4 096	780	3 120	12 480	194.96	0.002
2009	5	210	25	125	625	15 625	1 050	5 250	26 250	203.4	43.56
2010	6	225	36	216	1 296	46 656	1 350	8 100	48 600	222.24	7.62
$\sum$	0	2 610	182	0	4 550	134 342	510	34 380	21 300	2 609.98	2 836.29

【解】 (1) 绘制散点图,选择预测模型。

将时间序列观察期资料绘在坐标图上(见图 5-1),观察其发展趋势,呈现先上升,后下降,再上升的发展趋势,其图形为一条有两个弯曲的曲线,因此,可配合三次抛物线预测模型来预测。

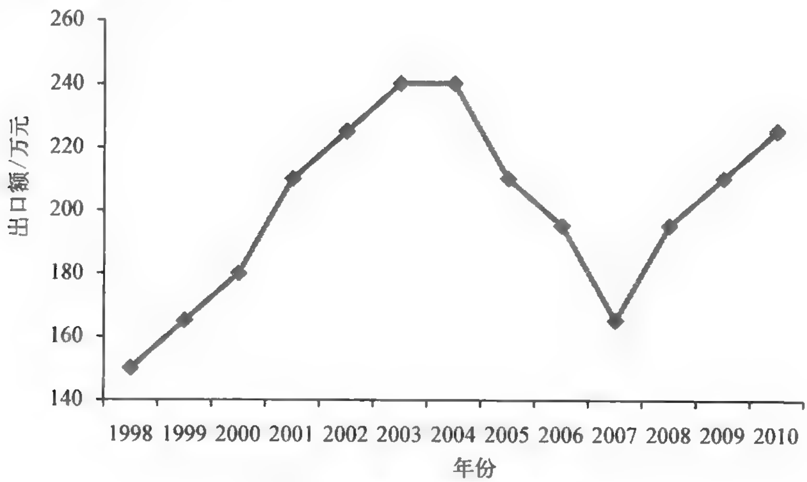


图 5-1 某公司商品出口额

(2) 建立三次抛物线预测模型。

列表计算求解与待定参数有关的数据资料，如表 5-4 所示。

将计算结果代入式(5-13)得

$$\begin{cases}2\ 160 = 13a + 182c \\ 510 = 182b + 4\ 550d \\ 34\ 380 = 182a + 4\ 550c \\ 21\ 300 = 4\ 550b + 134\ 342d\end{cases}$$

解此方程组，得

$$\hat{a} = 215.245\ 5,\ \hat{b} = -7.578$$
$$\hat{c} = -1.033\ 95,\ \hat{d} = 0.415\ 2$$

所求三次抛物线预测模型为

$$\hat{y}_t = 215.245\ 5 - 7.578t - 1.033\ 95t^2 + 0.415\ 2t^3$$

将各年的 t 值代入预测模型，可得各年的追溯预测值 $\hat{y}_t$。

2011 年的预测值为

$$\hat{y}_t = 215.245\ 5 - 7.578 \times 7 - 1.033\ 95 \times 7^2 + 0.415 \times 7^3 = 253.95\ (\text{万元})$$

计算标准误差：

$$S_y = \sqrt{\frac{\sum \alpha^{n-t}(y_t - \hat{y}_t)^2}{n-m}} = \sqrt{\frac{2\ 836.29}{13-4}} = 17.752\ 3$$

当 $a = 0.05$，自由度 $n - m = 9$，查 t 分布表得 t 临界值 $t_{0.025}(9) = 2.262$，则 2011 年商品出口额的区间估计为

$$\hat{y}_7 \pm t_{0.025}(9) S_y \sqrt{1 + \frac{1}{n}} = 253.95 \pm 2.262 \times 17.752\ 3 \times \sqrt{1 + \frac{1}{13}}$$
$$= 253.95 \pm 41.69$$

得预测区间估计值为(212.28，295.62)，即有 95%的把握预期 2011 年公司商品出口额将在(212.28～295.62)万元。

5.2.2 三点法

用三点法来估计二次抛物线预测模型的参数，其基本思想是：在二次抛物线上选取三个代表点来求模型的三个参数估计值。这三点选择方法是：当时间序列

的总项数 $n \geqslant 15$ 时，在序列的首尾两端和正中各取五项数据，求出三个加权平均数，权数由远及近分别用 1,2,3,4,5，用以加重近期信息在平均数中的比重。这三个加权平均数就作为二次抛物线上三个点的纵坐标。若 $9 \leqslant n \leqslant 15$ 时，则在序列初、中、近期各取三项求出三个加权平均数，权数由远及近分别用 1,2,3。为了保持这三个点之间的距离相等，数列总项数应为奇数，若是偶数，可删去最早期的一项。

设初、中、近期三点的坐标为 $M_1(t_1, R)$, $M_2(t_2, S)$, $M_3(t_3, T)$。

又设 n 为数列总项数，且为奇数，则正中项为 $d = \frac{n+1}{2}$。

设各项观察值为 $y_1, y_2, \cdots, y_d, \cdots, y_n$。于是，五项加权平均时，三个加权平均数为

$$R = \frac{1}{15}(y_1 + 2y_2 + 3y_3 + 4y_4 + 5y_5)$$

$$S = \frac{1}{15}(y_{d-2} + 2y_{d-1} + 3y_d + 4y_{d+1} + y_{d+2})$$

$$T = \frac{1}{15}(y_{n-4} + 2y_{n-3} + 3y_{n-2} + 4y_{n-1} + 5y_n)$$

这三点的横坐标也应取加权平均值，即

$$t_1 = \frac{1}{15}(1 + 2\times 2 + 3\times 3 + 4\times 4 + 5\times 5) = \frac{11}{3} = 3 + \frac{2}{3}$$

$$t_2 = d + \frac{2}{3} = \frac{n+1}{2} + \frac{2}{3} = \frac{3n+7}{6}$$

$$t_3 = (n-2) + \frac{2}{3} = n - \frac{4}{3}$$

于是，五项加权平均时，三点的坐标为

$$M_1\left(\frac{11}{3}, R\right), M_2\left(\frac{3n+7}{6}, S\right), M_3\left(n - \frac{4}{3}, T\right)。$$

同理可求得三项加权平均时三点的坐标为

$$M_1\left(\frac{7}{3}, R\right), M_2\left(\frac{3n+5}{6}, S\right), M_3\left(n - \frac{2}{3}, T\right)$$

$$R = \frac{1}{6}(y_1 + 2y_2 + 3y_3)$$

$$S=\frac{1}{6}(y_{d-1}+2y_d+3y_{d+1})$$

$$T=\frac{1}{6}(y_{n-2}+2y_{n-1}+3y_n)$$

求得三点坐标后，就可求二次抛物线预测模型的参数估计值。二次抛物线预测模型为 $\hat{y}_t=a+bt+ct^2$。二次抛物线上的三点必须满足这模型。因此，五项加权平均时，有：

$$\begin{cases} R=a+\dfrac{11}{3}b+\left(\dfrac{11}{3}\right)^2c \\ S=a+\dfrac{3n+7}{6}b+\left(\dfrac{3n+7}{6}\right)^2c \\ T=a+\left(n-\dfrac{4}{3}\right)b+\left(n-\dfrac{4}{3}\right)^2c \end{cases}$$

解此方程组，可得参数估计值为

$$\begin{cases} \hat{c}=\dfrac{2(R+T-2S)}{(n-5)^2} \\ \hat{b}=\dfrac{T-R}{n-5}-\dfrac{3n+7}{3}\hat{c} \\ \hat{a}=R-\dfrac{11}{3}\hat{b}-\dfrac{121}{9}\hat{c} \end{cases} \tag{5-14}$$

同理，三项加权平均时，可求得参数的估计值为

$$\begin{cases} \hat{c}=\dfrac{2(R+T-2S)}{(n-3)^2} \\ \hat{b}=\dfrac{T-R}{n-3}-\dfrac{3n+5}{3}\hat{c} \\ \hat{a}=R-\dfrac{7}{3}b-\dfrac{49}{9}c \end{cases} \tag{5-15}$$

【例 5-4】 观察期数据同上例，试用三点法建模，并预测该公司 2011 年的商品出口额。

【解】 (1) 选择预测模型。

计算序列的一阶、二阶差分，列于表 5-5 中，从计算结果可看出，二阶差分是比较平稳的。因此，选择二次抛物预测模型来预测。

(2) 列表计算有关数据,建立二次抛物线预测模型。

表 5-5 某公司商品销售收入及其差分 单位:万元

年 份	2002	2003	2004	2005	2006	2007	2008	2009	2010
销售收入 y_t	545	641	764	923	1 107	1 322	1 568	1 836	2 140
一阶差分 ∇y_t		96	123	159	184	215	246	268	304
二阶差分 $\nabla^2 y_t$			27	36	25	31	31	22	36

表 5-6 某公司商品出口额二次抛物线预测模型三点法计算表 单位:万元

年 份	t	销售收入 y_t	权数 ω	ωy_t	$\hat{y}_t$	$(y_t-\hat{y}_t)^2$
2002	1	545	1	545	549.68	21.90
2003	2	641	2	1 282	647.44	41.47
2004	3	764	3	2 292	774.34	106.92
2005	4	923	1	923	930.38	54.46
2006	5	1 107	2	2 214	1 115.56	73.27
2007	6	1 322	3	3 966	1 329.88	62.09
2008	7	1 568	1	1 568	1 573.34	28.52
2009	8	1 836	2	3 672	1 845.94	98.80
2010	9	2 140	3	6 420	2 147.68	58.98
$\sum$		—	—	—	—	546.41

根据上表资料计算得:其中 $d=\dfrac{n+1}{2}=5$ (n 为数列总项数)

$$R=\frac{1}{6}(y_1+2y_2+3y_3)=\frac{545+1\,282+2\,292}{6}=686.5$$

$$S=\frac{1}{6}(y_4+2y_5+3y_6)=\frac{923+2\,214+3\,966}{6}=1\,183.8$$

$$T=\frac{1}{6}(y_7+2y_8+3y_9)=\frac{1\,568+3\,672+6\,420}{6}=1\,943.3$$

代入式(5-14),可得

$$\hat{c}=\frac{2(R+T+2S)}{(n-3)^2}=\frac{2\times(686.5+1\,943.3-2\times1\,183.8)}{(9-3)^2}=14.57$$

$$\hat{b}=\frac{T-R}{n-3}-\frac{3n+5}{3}\hat{c}=\frac{1\,943.3-686.5}{9-3}-\frac{3\times 9+5}{3}\times 14.57=54.05$$

$$\hat{a}=R-\frac{7}{3}b-\frac{49}{9}c=686.5+\frac{7}{3}\times 54.05-\frac{49}{9}\times 14.57=481.06$$

用三点法建立的二次抛物线预测模型为

$$y_t=481.06+54.05t+14.57t^2$$

将各年的 t 值代入预测模型，可得各年的追溯预测值 $\hat{y}_t$，并填入表中。

(3) 预测。

先计算 S_y，将表 5-6 计算结果代入，便得

$$S_y=\sqrt{\frac{\sum a^{n-t}(y_t-\hat{y}_t)^2}{n-m}}=\sqrt{\frac{546.41}{9-3}}=9.54$$

又 $\alpha=0.05$，自由度 $n-m=6$ 时，查 t 分布表得

$$T_{0.025}(6)=2.447$$

将 $t=10$ 代入预测模型，可得 2011 年的预测值为

$$\hat{y}_{10}=481.06+54.05\times 10+14.57\times 10^2=2\,478.56(\text{万元})$$

预测区间为

$$\hat{y}_{10}\pm t_{\alpha/2}S_y\sqrt{1+\frac{1}{n}}=2\,478.56\pm 2.447\times 9.54\times\sqrt{1+\frac{1}{9}}=2\,478.56\pm 24.61$$

即预测区间为(2 453.95，2 503.17)。

运用三点法同样可以估计直线或三次抛物线预测模型的参数。但是，这时就要选取两个或四个代表点。这里不详细讨论了，所以三点法是广义的。

5.3 指数曲线模型预测法

技术发展、社会发展的大量定量特性表现为随时间按指数或接近指数规律增长，例如飞机速度、光源效率等，因此，利用指数趋势模型来进行外推预测在实际中具有很广泛的应用。

指数曲线预测模型为

$$\hat{y}_t=ab^t \tag{5-16}$$

式中，a、b 为参数；t 为时间。

指数曲线预测模型的特点，是环比发展速度为一常数，$\hat{y}_t/\hat{y}_{t-1}=b$，这就是说时间序列按相同的增长率增减变化，或是时间序列的逐期增长量是不断递增或递减的。因此，当时间序列 $\{y_t\}$ 的环比发展速度大体相等，或对数一阶差分近似为一常数时，可配合指数曲线预测模型来预测。

若对模型两边取对数，则可化为对数直线模型，即

$$\lg \hat{y}_t = \lg a + t\lg b \tag{5-17}$$

其特点是对数的一阶差分为一常数：

$$\nabla(\lg \hat{y}_t) = \lg b$$

可以看出，$\lg \hat{y}_t$ 依赖于时间 t 做线性变化，即在半对数的坐标图中，指数曲线转变为一条直线。因此，可以先将时间序列 y_t 取对数后，用变换后的新序列与时间 t 建立线性模型，从而可以利用线性模型的参数估计方法来求出曲线参数，可用最小平方法、三点法等来估计，然后通过 $\lg a$、$\lg b$ 的反对数求 a、b 的值。

5.3.1 最小平方法

根据最小平方法的原理，对于对数直线预测模型，可得标准方程组为

$$\begin{cases}\sum \lg y_t = n\lg a + \lg b\sum t \\ \sum t\lg y_t = \lg a\sum t + \lg b\sum t^2\end{cases} \tag{5-18}$$

若选取时间序列 $\{y_t\}$ 的中点为时间原点，可使 $\sum t = 0$，则上述方程组简化为

$$\begin{cases}\sum \lg y_t = n\lg a \\ \sum t\lg y_t = \lg b\sum t^2\end{cases}$$

由此可得

$$\lg \hat{a} = \frac{\sum \lg y_t}{n}$$

$$\lg \hat{b} = \frac{\sum t\lg y_t}{\sum t^2} \tag{5-19}$$

求反对数，便得 a、b 的估计值。

【例 5-5】 某航运企业 2007—2013 年全部在职人员总数如表 5-7 第(3)栏

所示，试用最小平方法求参数预测 2014 年在职人员总数。

【解】 列最小平方法参数计算表，如表 5－7 所示。

表 5－7 最小平方法计算表 单位：人

年 份	t	在职人员总数 y_t	$y'=\lg y$	t^2	ty'
(1)	(2)	(3)	(4)	(5)	(6)
2007	−3	2 497	3.397 4	9	−10.192 3
2008	−2	2 558	3.407 9	4	−6.815 8
2009	−1	2 624	3.419 0	1	−3.419 0
2010	0	2 671	3.426 7	0	0
2011	1	2 748	3.439 0	1	3.439 0
2012	2	2 778	3.443 7	4	6.887 4
2013	3	2 813	3.449 2	9	10.347 5
$\sum$	0		23.982 9		0.247 0

$$\lg\hat{a}=\frac{\sum \lg y_t}{n}=\frac{23.982\,9}{7}\approx 3.426\,1$$

$$\lg\hat{b}=\frac{\sum t\lg y_t}{\sum t^2}=\frac{0.247\,0}{28}\approx 0.008\,8$$

求反对数得

$$\hat{a}=2\,668.09,\ \hat{b}=1.02$$

所求指数曲线预测模型为

$$\hat{y}_t=\hat{a}\hat{b}^t=2\,668.09\times(1.02)^t$$

令 $t=4$，代入预测模型，可得 2014 年该航运企业在职人数的预测值为

$$\hat{y}_{2008}=2\,668.09\times(1.02)^4\approx 2\,888(\text{人})$$

5.3.2 三点法

三点法也可用于估计指数曲线预测模型的参数，但是，这时应只选两个代表取点。指数曲线的对数形式为

$$\lg \hat{y}_t = \lg a + t\lg b$$

令
$$\hat{y}'_t = \lg \hat{y}_t,\ a' = \lg a,\ b' = \lg b$$

则上述就化为直线预测模型：

$$\hat{y}'_t = a' + b't$$

对于直线预测模型，类似前节介绍的三点法，若 $n > 10$，在序列首尾两端各取五项加权平均，便可推得参数估计值为

$$\hat{b}' = \frac{T-R}{n-5}$$

$$\hat{a}' = R - \frac{11}{3}\hat{b}' \tag{5-20}$$

若 $n < 10$，则取三项加权平均，同理可求得参数估计值为

$$\hat{b}' = \frac{T-R}{n-3}$$

$$\hat{a}' = R - \frac{7}{3}\hat{b}' \tag{5-21}$$

【例 5-6】 根据表 5-8 某港口企业流动资金总额的统计资料，试用三点法建立预测模型，预测 2013 年该企业流动资金总额。

表 5-8 某港口企业流动资金总额指数曲线预测模型三点法计算表 单位：百万元

年份	t	流动资金额 y_t	$y' = \lg y$	权数 ω	$\omega y'$	$\hat{y}_t$
2001	1	5.67	0.753 58	1	0.753 58	5.379 1
2002	2	7.09	0.850 65	2	1.701 30	7.181 7
2003	3	9.56	0.980 46	3	2.941 38	9.588 2
2004	4	13.07	1.116 28	4	4.465 12	12.801 3
2005	5	16.75	1.224 01	5	6.120 05	17.091 0
2006	6	21.62	1.334 86	小计	15.981 431	22.818 1
2007	7	28.34	1.452 40	1	1.600 54	30.464 5
2008	8	39.86	1.600 54	2	3.467 36	40.673 1
2009	9	54.16	1.733 68	3	5.622 39	54.302 7

续 表

年份	t	流动资金额 y_t	$y'=\lg y$	权数 ω	$\omega y'$	$\hat{y}_t$
2010	10	74.84	1.874 13	4	7.899 52	72.499 5
2011	11	94.38	1.974 88	5	10.568 70	96.794 1
2012	12	129.94	2.113 74			129.229 8
				小计	29.158 51	

【解】 (1) 建立指数曲线预测模型。

根据表 5-8 的资料计算得

$$R=\frac{15.981\ 43}{15}=1.065\ 4$$

$$T=\frac{29.158\ 51}{15}=1.943\ 9$$

代入式(5-20),可得

$$\hat{b}'=\frac{1.943\ 9-1.065\ 4}{12-5}=0.125\ 5$$

$$\hat{a}'=1.065\ 4-\frac{11}{3}\times 0.125\ 5=0.605\ 2$$

于是所求预测模型为

$$\hat{y}'_t=0.605\ 2+0.125\ 5t$$

而

$$\hat{a}'=\lg\hat{a}=0.605\ 2$$
$$\hat{b}'=\lg\hat{b}=0.125\ 5$$

求反对数得

$$\hat{a}=4.029\ 0,\ \hat{b}=1.335\ 1$$

因此,所求指数曲线预测模型为

$$\hat{y}_t=4.029\ 0\times(1.335\ 1)^t$$

将各年的 t 值代入预测模型,可得各年追溯预测值 $\hat{y}_t$,如表 5-8 所示。

(2) 预测。

以 $t=13$ 代入预测模型，可得 2013 年港口企业流动资金总额的预测值为

$$\hat{y}_{2013}=4.0290\times(1.3351)^t=172.5348\text{（百万元）}$$

若把这两种方法的预测结果比较一下，显然，三点法的预测结果要比最小平方法的大。这是因为流动资金逐年增加，且 2012 年增加得特别多，而三点法对远、近期水平却采用了不同权数进行加权的原因。

5.4 修正指数曲线模型预测法

某些经济现象在其发展过程中，通常都有这样的现象，即初期增长速度较快，随后增长速度减慢，逐渐达到某一稳定状态。这种发展趋势可以用修正指数趋势模型来描述。修正指数曲线是一种渐近增长曲线，它在工业生产的需求预测中具有一定的作用。

5.4.1 预测模型及其特征

修正指数曲线预测模型为

$$\hat{y}_t=k+ab^t$$

式中，k、a、b 为参数；t 为时间。

修正指数曲线模型只比指数曲线模型多一个 k $(k>0)$ 值，它是对指数曲线模型的某种修正。

求 $\hat{y}_t$ 的一阶、二阶导数可得

$$\hat{y}_t'=(a\ln b)b^t,\ \hat{y}_t''=a(\ln b)^2b^t$$

(1) 当 $k>0$, $a<0$, $0<b<1$ 时，有 $\hat{y}_t'>0$, $\hat{y}_t''<0$；此时 $\hat{y}_t$ 是单调递增的，$\hat{y}_t'$ 是递减的，可知 $\hat{y}_t'$ 的图形是凸的。当 $t=0$ 时，$\hat{y}_t=k+a(a<0)$，当 $t\to+\infty$ 时，因为 $b^t\to0$，可得 $\hat{y}_t\to k$，因此 $\hat{y}_t=k$ 是它的渐近线。由以上分析可知，$\hat{y}_t$ 随着 t 的增加而增加，增长速度是先快后慢，最后接近于高限 k。其图形如图 5-2 所示。

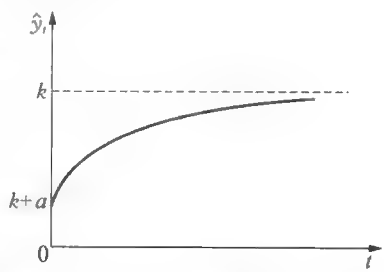


图 5-2 $\hat{y}_t=k+ab^t(k>0,\ a<0,\ 0<b<1)$

因此，当某变量的变动规律是初期增长较快，随后增长速度逐渐放慢，最后趋向某一正常极限时，可用修正指数曲线来描述。

(2) 当 $k>0$, $a>0$, $0<b<1$ 时，有 $\hat{y}_t'<0$, $\hat{y}_t''>0$；此时 $\hat{y}_t$ 是递减的，且图形是凹的。当 $t=0$ 时，$\hat{y}_t=k+a$ $(a>0)$，当 $t\to+\infty$ 时，可得 $\hat{y}_t\to k$ 是它的渐近线。这说明 $\hat{y}_t$ 随着 t 的增加而减少，递减速度是先快后慢，最后接近于底限 k，其图形如图 5-3 所示。

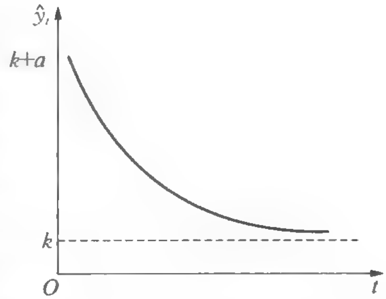


图 5-3 $\hat{y}_t=k+ab^t(k>0,\ a>0,\ 0<b<1)$

由上述分析可得：修正指数曲线还可用来描述初期减少较快，随后减少比较缓慢，最后趋向某一正常数极限的经济变量。

由于修正指数曲线预测模型的一阶差分

$$\begin{aligned}\nabla\hat{y}_t&=(k+ab^t)-(k+ab^{t-1})\\&=a(b-1)b^{t-1}\end{aligned}$$

是指数函数形式，因此由指数曲线预测模型的特点，可知修正指数曲线预测模型的特征是：一阶差分的环比为一常数。

当时间序列$\{y_t\}$的一阶差分为 Δy_t 的环比近似一常数时，可配合修正指数曲线预测模型来预测。

5.4.2 预测模型参数估计方法

由于修正指数曲线模型比较复杂，不宜用最小平方法求其参数。参数 k、a、b 的估计，通常采用三段法和三点法。这里介绍三段法。

设有 N $(N=3n,\ n\geqslant2)$ 个历史数据如表 5-9 所示。

表 5-9 三段法历史数据表

t	0	1	…	$n-1$	…	$2n-1$	…	$3n-1$
y_t	y_0	y_1	…	y_{n-1}	…	y_{2n-1}	…	y_{3n-1}

如果通过分析，所有历史数据都近似在修正指数曲线方程 $\hat{y}_t=k+ab^t(t=0,1,\cdots,3n-1)$ 上，则序列$\{y_t\}$的发展趋势可用修正指数曲线来描述。

把序列$\{y_t\}$平均分为三段，每段含有 n 个数据，对各段求和，可得

$$\sum\nolimits_1 y_t=\sum_{t=0}^{n-1}y_t=nk+a(b^0+b^1+\cdots+b^{n-1})$$

$$= nk + a\frac{b^n - 1}{b - 1} \tag{5-22}$$

$$\sum\nolimits_2 y_t = \sum_{t=n}^{2n-1} y_t = nk + ab^n(b^0 + b^1 + \cdots + b^{n-1})$$

$$= nk + ab^n \frac{b^n - 1}{b - 1} \tag{5-23}$$

$$\sum\nolimits_3 y_t = \sum_{t=2n}^{3n-1} y_t = nk + ab^{2n}(b^0 + b^1 + \cdots + b^{n-1})$$

$$= nk + ab^{2n} \frac{b^n - 1}{b - 1} \tag{5-24}$$

于是

$$\sum\nolimits_2 y_t - \sum\nolimits_1 y_t = a\frac{(b^n - 1)^2}{b - 1} \tag{5-25}$$

$$\sum\nolimits_3 y_t - \sum\nolimits_2 y_t = ab^n \frac{(b^n - 1)^2}{b - 1} \tag{5-26}$$

从而有

$$b^n = \frac{\sum\nolimits_3 y_t - \sum\nolimits_2 y_t}{\sum\nolimits_2 y_t - \sum\nolimits_1 y_t}$$

即

$$\hat{b} = \sqrt[n]{\frac{\sum\nolimits_3 y_t - \sum\nolimits_2 y_t}{\sum\nolimits_2 y_t - \sum\nolimits_1 y_t}} \tag{5-27}$$

由式(5-25)可得

$$\hat{a} = \left(\sum\nolimits_2 y_t - \sum\nolimits_1 y_t\right)\frac{\hat{b} - 1}{(\hat{b}^n - 1)^2} \tag{5-28}$$

由式(5-22)可得

$$\hat{k} = \frac{1}{n}\left[\sum\nolimits_1 y_t - \hat{a}\left(\frac{\hat{b}^n - 1}{\hat{b} - 1}\right)\right] \tag{5-29}$$

【例 5-7】 假设某港口 2002—2007 年的货物吞吐量如表 5-10 所示，试预测

2008 年和 2009 年的吞吐量。

表 5-10　某港吞吐量修正指数曲线预测模型计算表　　单位：百万吨

年　份	t	实际吞吐量 y_t	$\sum y_t$
2002	0	20.81	$\sum_1 y_t = 47.92$
2003	1	27.11	
2004	2	30.33	$\sum_2 y_t = 64.68$
2005	3	34.35	
2006	4	37.91	$\sum_3 y_t = 79.95$
2007	5	42.04	

【解】 列表计算有关数据可见表 5-10。$n = \dfrac{6}{3} = 2$。将计算结果代入 k、a、b 的计算公式，可得

$$b = \left[\frac{\sum_3 y_t - \sum_2 y_t}{\sum_2 y_t - \sum_1 y_t}\right]^{\frac{1}{n}} = \left(\frac{79.95 - 64.68}{64.68 - 47.92}\right)^{\frac{1}{2}} \approx 0.955$$

$$a = \left(\sum_2 y_t - \sum_1 y_t\right)\frac{b-1}{(b^n - 1)^2} = (64.68 - 47.92) \times \frac{0.955 - 1}{(0.955^2 - 1)^2} = -94.275$$

$$k = \frac{1}{n}\left[\frac{\sum_1 y_t \sum_3 y_t - \left(\sum_2 y_t\right)^2}{\sum_1 y_t + \sum_3 y_t - 2\sum_2 y_t}\right] = \frac{1}{2}\left(\frac{47.92 \times 79.95 - 64.68^2}{47.92 + 79.95 - 2 \times 64.68}\right) = 118.221$$

可得修正指数曲线预测模型为

$$\hat{y}_t = k + ab^t = 118.221 + (-94.275) \times (0.995)^t$$

将各年的 t 值代入预测模型，可得各年追溯预测值。令 $t = 6$，$t = 7$，可得 2008 年和 2009 年的预测值：

$$\hat{y}_{2008} = 118.221 + (-94.275) \times (0.995)^6 = 46.70(\text{百万吨})$$

$$\hat{y}_{2009} = 118.221 + (-94.275) \times (0.995)^7 = 49.92(\text{百万吨})$$

因为 $k = 118.22 > 0$，$a = -94.275 < 0$，$0 < b = 0.955 < 1$，可推知该港口正处在生产周期的成长期阶段。2008 年、2009 年的预测值离 k 值（渐近线）还有相当距离，该港口可结合实际经济背景，积极开拓腹地货源，进一步提高港口的吞吐量。

5.5 成长曲线预测

成长曲线模型是指龚柏兹(Gompertz)曲线模型和罗吉斯蒂(Logistic)曲线模型。

5.5.1 龚柏兹曲线预测模型

龚柏兹曲线，是美国统计学家和数学家龚柏兹首先提出用作控制人口增长率的一种数学模型，随后被广泛应用于耐用消费品的产品生命周期预测。他的预测模型为

$$\hat{y}_t = ka^{b^t}\ (k > 0) \tag{5-30}$$

式中，k、a、b 为参数；t 为时间。

下面分析该模型的图形：对 $\hat{y}_t$ 求一阶、二阶导数，有

$$\hat{y}_t' = ka^{b^t} b^t \ln a \ln b$$

$$\hat{y}_t'' = ka^{b^t} b^t \ln a (\ln b^2)(b^t \ln a + 1)$$

令 $\hat{y}_t'' = 0$，可求得曲线拐点的位置为

$$\left(\frac{\ln[-(\ln a)^{-1}]}{\ln b}, \frac{k}{e}\right) \quad 0 < a < 1$$

曲线过此点由向上凹变为向下凹。

当 $k > 0$，$0 < a < 1$，$0 < b < 1$ 时，由于 $\ln a < 0$，$\ln b < 0$，所以 $\hat{y}_t' > 0$，可知 $\hat{y}_t$ 为增函数，且在点 $\left(\frac{\ln[-(\ln a)^{-1}]}{\ln b}, \frac{k}{e}\right)$ 出现转折，$\hat{y}_t$ 的增长率由逐渐增大变为逐渐减小。在 $t = 0$ 时，$\hat{y}_t = ka$。当$t \to -\infty$ 时，由于 $b^t \to \infty$，$a^{b^t} \to 0$，有 $\hat{y}_t \to 0$；当 $t \to +\infty$ 时，由于 $b^t \to 0$，$a^{b^t} \to 1$，有 $\hat{y}_t \to k$。所以 $\hat{y}_t = 0$ 和 $\hat{y}_t = k$ 都是它的渐近线。它的图形是一条 S 型曲线(见图5-4)。这条曲线反映了某些经济变量由开始增长缓慢，随后增长加快，达到一定程度后，增长率逐渐减慢，最后达到饱和状态的过

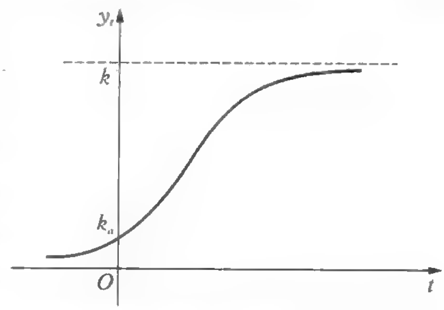


图 5-4 $\hat{y}_t = ka^{b^t}$ ($k > 0$, $0 < a < 1$, $0 < b < 1$)

程。因此,对于具有这种发展趋势的预测目标,可考虑用龚柏兹曲线来描述。

为了确定模型中的参数,通常把预测模型改写为对数形式:

$$\lg \hat{y}_t = \lg k + (\lg a)b^t \tag{5-31}$$

若分别以 $\hat{y}_t$、K、A 取代 $\lg \hat{y}_t$、$\lg k$、$\lg a$,则上式转化为修正指数曲线预测模型:

$$\hat{y}_t = K + Ab^t$$

因此可以用修正指数曲线估计参数的方法,利用三段法求得参数 K、A、b。

$$\hat{b} = \sqrt[n]{\frac{\sum_3 \lg y_t - \sum_2 \lg y_t}{\sum_2 \lg y_t - \sum_1 \lg y_t}}$$

$$\ln \hat{a} = \left(\sum_2 \lg y_t - \sum_1 \lg y_t\right) \frac{\hat{b} - 1}{(\hat{b}_n - 1)^2}$$

$$\lg \hat{k} = \frac{1}{n}\left[\sum_1 \lg y_t - \left(\frac{\hat{b}^n - 1}{\hat{b} - 1}\right) \lg \hat{a}\right] \tag{5-32}$$

这里 n 为总数据的 1/3。$\sum_1 \lg y_t$、$\sum_2 \lg y_t$、$\sum_3 \lg y_t$ 分别为总数据三等分后的各部分和。求 $\lg a$、$\lg k$ 的反对数即得参数 A、K。

由于龚柏兹曲线的对数形式为修正指数曲线,因而根据修正指数曲线预测模型的特点,可知龚柏兹曲线预测模型的特征是,其对数一阶差分的环比为一常数。因此,当时间序列$\{y_t\}$的对数一阶差分的环比近似一常数时,可配合龚柏兹曲线预测模型来预测。

【例 5-8】 某省 2002—2007 年年底电冰箱社会拥有量如表 5-9 所示,试用龚柏兹曲线预测模型来预测 2008 年和 2009 年年底该省电冰箱社会拥有量。

【解】 建立龚柏兹曲线预测模型计算有关数据,将历年实际值转化为对数形式。如表 5-11 所示。

表 5-11 三段法求参数计算表

年　份	t	实际值 y_t	$\lg y_t$
2002	0	410	2.612 8
2003	1	727	2.861 5
2004	2	1 181	3.072 2
2005	3	1 927	3.284 9
2006	4	2 554	3.407 2
2007	5	2 996	3.476 5

$$\sum\nolimits_1 \lg y_t = 2.6128 + 2.8615 = 5.4743$$
$$\sum\nolimits_2 \lg y_t = 3.0722 + 3.2849 = 6.3571$$
$$\sum\nolimits_3 \lg y_t = 3.4072 + 3.4765 = 6.8837$$

将计算结果代入 b、$\lg a$ 和 $\lg k$ 的计算公式。可得

$$b = \sqrt{\frac{\sum_3 \lg y_t - \sum_2 \lg y_t}{\sum_2 \lg y_t - \sum_1 \lg y_t}} = \left(\frac{6.8837 - 6.3571}{6.3571 - 5.4743}\right)^{\frac{1}{2}} \approx 0.7723$$

$$\lg a = \left(\sum\nolimits_2 \lg y_t - \sum\nolimits_1 \lg y_t\right)\frac{b-1}{(b^2-1)^2} = 0.8828 \times \frac{-0.2277}{(0.5965-1)^2} \approx -1.2347$$

求反对数，可得

$$a = 0.0583$$

$$\begin{aligned}\lg k &= \frac{1}{n}\left[\frac{\sum_1 \lg y_t \sum_3 \lg y_t - \left(\sum_2 \lg y_t\right)^2}{\sum_1 \lg y_t + \sum_3 \lg y_t - 2\sum_2 \lg y_t}\right] \\ &= \frac{1}{2} \times \left(\frac{5.4743 \times 6.8837 - 6.3571^2}{5.4743 + 6.8837 - 2 \times 6.3571}\right) \\ &\approx 3.8311\end{aligned}$$

求反对数，可得

$$K = 6\,777.9756$$

于是，所求龚柏兹曲线预测模型为

$$\hat{y}_t = ka^{b^t} = 6\,777.9756 \times (0.0583)^{(0.7723)^t}$$

将各年的 t 值代入预测模型，可得各年追溯预测值。

由于用此模型预测计算太繁琐，所以可用取对数后的模型：

$$\lg \hat{y}_t = \lg k + b^t \lg a = 3.8311 + (-1.2347) \times (0.7723)^t$$

将 $t = 6$ 和 $t = 7$ 分别代入预测模型求反对数后，可得 2008 年和 2009 年的预测值分别为

$$\lg \hat{y}_{2008} = 3.8311 + (-1.2347) \times (0.7723)^6 = 3.5691$$
$$\lg \hat{y}_{2009} = 3.8311 + (-1.2347) \times (0.7723)^7 = 3.6288$$

可得

$$\hat{y}_{2008} \approx 3\ 707.660\ 8(\text{千台})$$
$$\hat{y}_{2009} = 4\ 254.024\ 6(\text{千台})$$

因为 $\lg a = -1.234\ 7 < 0$，$0 < b = 0.772\ 3 < 1$。

从预测结果可看出，该省 2002—2007 年电冰箱社会拥有量以较快的速度增长，距饱和水平 k ($k = 6\ 777.975\ 6$ 千台) 还有一定距离，市场需求仍较大。

5.5.2 罗吉斯缔曲线预测模型

罗吉斯缔曲线(又称逻辑曲线或推理曲线)，它是由比利时数学家维哈尔斯特(Veihulot)在研究人口增长规律时提出来的。它又称为皮尔生长理论曲线。

罗吉斯缔曲线预测模型为

$$\hat{y}_t = \frac{1}{k + ab^t} \tag{5-33}$$

或

$$\frac{1}{\hat{y}_t} = k + ab^t$$

式中，k、a、b 为参数；t 为时间。

对 $\hat{y}_t$ 求一阶、二阶导数，有并令 $\hat{y}''_t = 0$，可求得曲线拐点的位置为

$$\hat{y}'_t = \frac{-ab^t \ln b}{(k + ab^t)^2}$$

$$\hat{y}''_t = -\frac{a(\ln b)^2 b^t (k - ab^t)}{(k + ab^t)^3}$$

并令 $\hat{y}''_t = 0$，可求得曲线拐点的位置为

$$\left(\frac{\ln k - \ln a}{\ln b}, \frac{1}{2k}\right)$$

曲线过此点由向上凹变为向下凹。

当 $k > 0$，$a > 0$，$0 < b < 1$ 时，由于 $\ln b < 0$，所以 $\hat{y}'_t > 0$，此时 $\hat{y}_t$ 为增函数，且在 $\left(\frac{\ln k - \ln a}{\ln b}, \frac{1}{2k}\right)$ 出现转折，$\hat{y}_t$ 的增长率由逐渐增大变为逐渐减小。在 $t = 0$ 时，$\hat{y}'_t = \frac{1}{k + a}$。

当 $t \to -\infty$ 时，$\hat{y}_t \to 0$；当 $t \to \infty$ 时，$\hat{y}_t \to \frac{1}{k}$，所以 $\hat{y}_t = 0$ 和 $\hat{y}_t = \frac{1}{k}$ 都是它的渐近线，它的图形也是一条 S 形曲线，且对于拐点是对称的(见图 5-5)。它与龚柏

兹曲线很相似，也是描述某些变量由开始增长缓慢，随后增长加快，达到一定程度后，增长率较低直至平稳发展。因此，对于具有这种发展趋势的预测目标，根据其具体情况，可考虑用罗吉斯缔曲线描述。它能被用于产品生命周期中投入期、成长期和成熟前期的预测。

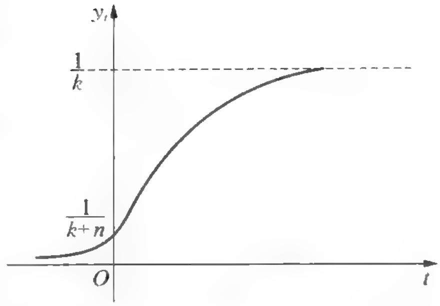


图 5-5 $\hat{y}_t=\dfrac{1}{k+ab^t}$ $(k>0,\ a>0,\ 0<b<1)$

由于罗吉斯缔曲线的倒数是修正指数曲线。因此，仿照修正指数曲线估计参数的方法，可得 b、a 和 k 的计算公式：

$$\hat{b}=\sqrt[n]{\frac{\sum_3 \frac{1}{y_t}-\sum_2 \frac{1}{y_t}}{\sum_2 \frac{1}{y_t}-\sum_1 \frac{1}{y_t}}}$$

$$\hat{a}=\left(\sum_2 \frac{1}{y_t}-\sum_1 \frac{1}{y_t}\right)\frac{\hat{b}-1}{(\hat{b}^n-1)^2}$$

$$\hat{k}=\frac{1}{n}\left[\frac{\sum_1 \frac{1}{y_t}\sum_3 \frac{1}{y_t}-\left(\sum_2 \frac{1}{y_t}\right)^2}{\sum_1 \frac{1}{y_t}+\sum_3 \frac{1}{y_t}-2\sum_2 \frac{1}{y_t}}\right]$$

或

$$\hat{k}=\frac{1}{n}\left(\sum_1 \frac{1}{y_t}-a\frac{b^n-1}{b-1}\right)$$

这里 n 为总数据的 1/3，$\sum_1 \frac{1}{y_t}$、$\sum_2 \frac{1}{y_t}$ 和 $\sum_3 \frac{1}{y_t}$ 分别为总数据三等分后的各部分和。若 $1/y_t$ 为小数时，可乘以 10 的适当乘方化为整数，以利计算。

同时根据修正指数曲线预测模型的特点，可知罗吉斯缔曲线预测模型的特征是其倒数一阶差分的环比为一常数。因此，它适用于历史数据取倒数后的一阶增长量环比系数比较接近的预测对象。

【例 5-9】 某省 2002—2007 年平均每百户农民家庭洗衣机拥有量如表 5-12 所示，试用罗吉斯缔曲线模型预测法预测 2008 年和 2009 年洗衣机拥有量。

【解】 建立罗吉斯缔曲线预测模型列表计算有关数据，如表 5-12 所示。

表 5-12　三段法求参数计算表　单位：台

年　份	t	实际值 y_t	$\frac{1}{y_t}$	$\sum \frac{1}{y_t}$
2002 2003	0 1	0.11 0.44	9.090 9 2.272 7	$\sum_1 \frac{1}{y_t} = 11.363\,6$
2004 2005	2 3	0.78 2.55	1.287 1 0.392 2	$\sum_2 \frac{1}{y_t} = 1.674\,3$
2006 2007	4 5	3.82 5.59	0.261 8 0.178 9	$\sum_3 \frac{1}{y_t} = 0.440\,7$

将计算结果代入 k、a、b 的计算公式，可得

$$b = \sqrt{\frac{\sum_3 \frac{1}{y_t} - \sum_2 \frac{1}{y_t}}{\sum_2 \frac{1}{y_t} - \sum_1 \frac{1}{y_t}}} = \sqrt{\frac{0.440\,7 - 1.674\,3}{1.674\,3 - 11.363\,6}} \approx 0.356\,8$$

$$a = \left(\sum_2 \frac{1}{y_t} - \sum_1 \frac{1}{y_t}\right)\frac{b-1}{(b^2-1)^2} = (-9.689\,3) \times \frac{-0.643\,2}{0.761\,6} = 8.183\,0$$

$$k = \frac{1}{2}\left[\frac{\sum_1 \frac{1}{y_t}\sum_3 \frac{1}{y_t} - \left(\sum_2 \frac{1}{y_t}\right)^2}{\sum_1 \frac{1}{y_t} + \sum_3 \frac{1}{y_t} - 2\sum_2 \frac{1}{y_t}}\right]$$

$$= \frac{1}{2} \times \frac{11.363\,6 \times 0.440\,7 - 1.674\,3^2}{11.363\,6 + 0.440\,7 - 2 \times 1.674\,3} \approx 0.130\,4$$

于是，所求罗吉斯蒂曲线预测模型为

$$\frac{1}{\hat{y}_t} = k + ab^t = 0.130\,4 + 8.183\,0 \times (0.356\,8)^t$$

令 $t = 6$，$t = 7$ 代入，求得 2008 年和 2009 年的预测值：

$$\frac{1}{\hat{y}_{2008}} = 0.130\,4 + 8.183\,0 \times (0.356\,8)^6 = 0.147\,3$$

$$\hat{y}_{2008} = 1 \div 0.147\,3 \approx 6.79(\text{台})$$

$$\frac{1}{\hat{y}_{2009}} = 0.130\,4 + 8.183\,0 \times (0.356\,8)^7 = 0.136\,4$$

$$\hat{y}_{2009} = 1 \div 0.136\,4 \approx 7.33(\text{台})$$

思考与练习

1. 某物流公司 2000—2008 年的收入总额如题表 5-1 所示。

题表 5-1 单位：万元

年份	2004	2005	2006	2007	2008	2009	2010	2011	2012
收入额	52	54	58	61	64	67	71	74	77

试选择合适的预测模型，分别用最小平方法和折扣最小平方法估计参数，预测 2013 年和 2014 年的销售额及预测区间（$\alpha = 0.05$）。

2. 某省 1991—2003 年的进出口货物数量资料如题表 5-2 所示：

题表 5-2 单位：万吨

年　份	2001	2002	2003	2004	2005	2006	2007
货物数量	11.99	13.73	14.88	14.26	15.48	14.61	18.98
年　份	2008	2009	2010	2011	2012	2013	
货物数量	28.46	32.63	45.03	52.73	51.70	54.07	

试配合二次抛物线预测模型，用三点法估计参数，预测 2014—2020 年该省进出口货物数量。

3. 某地区 1983—2012 年的人口数如题表 5-3 所示。

题表 5-3 单位：万人

年　份	人口数	年　份	人口数	年　份	人口数
2001	3 232	1989	3 762	2007	4 492
2002	3 181	1990	3 940	2008	4 558
2003	3 286	1991	4 073	2009	4 628
2004	3 407	1992	4 199	2010	4 713
2005	3 523	1993	4 317	2011	4 803
2006	3 641	1994	4 409	2012	4 893

试选择合适的预测模型，预测该省 2013—2022 年的人口数。

6 马尔柯夫预测法

6.1 马尔柯夫预测法概述

马尔柯夫预测法是俄国数学家马尔柯夫于 1907 年提出,后由蒙特卡罗加以发展而建立的一种预测方法,它将时间序列看作是一个随机过程,通过对事物不同状态的初始概率与状态之间转移概率的研究,确定事物未来状态的变化趋势。目前,这种预测方法在人才、产品销售状态、市场占有率、产品期望利润、设备更新、新产品开发和证券投资等方面被广泛运用。

6.1.1 马尔柯夫过程概述

随机过程 $X(t)$是随时间 t 变化的随机变量。对于某一时刻 t_1,$X(t_1)$便是随机过程 $X(t)$在 $t=t_1$ 时的状态,这个状态是随机变量。在时刻 t_0 所处的状态为已知的条件下,随机过程在时刻 $t>t_0$ 时所处的状态仅与 t_0 时的状态有关,而与 t_0 以前的状态无关,则称该随机过程为马尔柯夫过程。用分布函数来描述,可表现为:若在条件 $X(t_i)=x_i(i=1, 2, \cdots, n-1)$ 下,$X(t_n)$的分布函数恰好等于在条件 $X(t_{n-1})=x_{n-1}$ 下 $X(t_n)$的分布函数,即

$$F(x_n;\ t_n \mid x_{n-1},\ x_{n-2},\ \cdots,\ x_1;\ t_{n-1},\ t_{n-2},\ \cdots,\ t_1) = F(x_n,\ t_n \mid x_{n-1},\ t_{n-1}) \tag{6-1}$$

则称 $X(t)$为马尔柯夫过程。式(6-1)说明 t_n 时的状态 x_n 仅与 t_{n-1}时的状态有关,而与 t_{n-1}以前的状态无关,这种性质称为马尔柯夫过程的无后效性。在较长时间后,马尔柯夫过程趋于稳定状态,且与其初始状态无关,这种性质称为马尔柯夫过程的遍历性。

6.1.2 马尔柯夫预测法的基本概念

6.1.2.1 马尔柯夫链

若随机过程的分布函数式(6-1)中的参数为非负整数,且具有马尔柯夫特性,则该随机过程为马尔柯夫链。也就是说,时间和状态都是离散量的马尔柯夫过程称为马尔柯夫链。

6.1.2.2 时期

时期是指研究对象所处时间的计量单位。若从 2010 年开始，则 2010 年为第一期。

6.1.2.3 状态

状态是指事物可能出现或存在的状况，即研究对象的具体表现。例如在港口吞吐量的预测中，状态为各不同经济腹地吞吐量特征（如货种、流量、流向等）的划分。采用不同的预测状态将会得到不同的预测结果。通常，对预测对象的状态划分有两种方法：一种是根据预测对象本身状态的界限划分，如机器状态有正常和异常之分；天气有晴、阴、雨状态之分；农业有丰收、平收、歉收之分等。另一类是根据实际情况人为划分，如产品在市场上畅销或滞销状态的划分可以根据获利多少进行，这种划分不仅因产品不同而异，而且还与预测者的偏好、素质有关。应注意同一事物的不同状态之间应相互独立，即事物不能同时存在于两种状态。

6.1.2.4 转移概率

客观事物可能有 $E_1, E_2, \cdots, E_n$ 共 n 种状态，且每次只能出现某一种状态，每一状态都具有 n 个转移方向（包括转向本身），即

$$E_i \to E_1,\ E_i \to E_2,\ \cdots,\ E_i \to E_i,\ \cdots,\ E_i \to E_n$$

将这种转移的可能性用概率描述就是状态转移概率。状态转移概率中最基本的是一步转移概率 $p(E_{m+1}=j \mid E_m=i)$，它表示由状态 E_m 经过一步转移到状态 E_{m+1} 的概率，记为 p_{ij}，即

$$p_{ij} = p(E_{m+1}=j \mid E_m=i) \quad (i,\ j=1,\ 2,\ \cdots,\ n) \tag{6-2}$$

马尔柯夫链的一步转移概率 p_{ij} 具有以下性质：

(1) $0 \leqslant p_{ij} \leqslant 1,\ i,\ j=1,\ 2,\ \cdots,\ n$。

(2) $\sum_{j=1}^{n} p_{ij} = 1,\ i,\ j=1,\ 2,\ \cdots,\ n$。

6.1.2.5 状态转移矩阵

设由状态 E_i 经过一个时期转移到状态 E_j 的一步转移概率为 p_{ij}，将所有的一步转移概率排成一个矩阵，即

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \tag{6-3}$$

此矩阵即为马尔柯夫状态转移矩阵。

【例 6-1】 码头作业有繁忙和停滞两种状态，记繁忙为状态①，停滞为状态②。表 6-1 所示是某码头的作业记录。试确定其状态转移矩阵 $\boldsymbol{P}$。

表 6-1 某码头的作业记录表

季 数	1	2	3	4	5	6	7	8	9	10	11	12
销售状态	繁①	繁①	滞②	繁①	滞②	滞②	繁①	繁①	繁①	滞②	繁①	滞②
季数	13	14	15	16	17	18	19	20	21	22	23	24
销售状态	繁①	繁①	滞②	滞②	繁①	繁①	滞②	繁①	滞②	繁①	繁①	繁①

由表 6-1 可知，从繁忙状态出发的 14 个状态转移中，有 7 个是从繁忙转入繁忙的。它们分别是：1→2，7→8，8→9，13→14，17→18，22→23，23→24。因此，由式(6-2)可得 $p_{11}=7/14=0.5$。

同样，表 6-1 中上述 14 个从繁忙状态出发的状态转移中，有 7 个是从繁忙转入停滞的。它们分别是：2→3，4→5，9→10，11→12，14→15，18→19，20→21。

同样，表 6-1 中，有 9 个是从停滞出发的状态转移，其中由停滞转入停滞的状态有 2 个。它们分别是：5→6，15→16。由此，$p_{22}=2/9=0.222$。

上述 9 个从滞销出发的状态转移中，有 7 个从停滞转入繁忙。它们分别是：3→4，6→7，10→11，12→13，16→17，19→20，21→22。因此 $p_{21}=7/9=0.778$。

由式(6-3)可得到该商品的销售状态转移矩阵 $\boldsymbol{P}$ 为

$$\boldsymbol{P}=\begin{bmatrix}p_{11} & p_{12}\\ p_{21} & p_{22}\end{bmatrix}=\begin{bmatrix}0.5 & 0.5\\ 0.778 & 0.222\end{bmatrix}$$

6.1.3 概率矩阵

为正确理解、熟练掌握、灵活运用马尔柯夫预测模型，下面简要介绍概率矩阵的概念与性质。

1) 概率向量

诸元素皆为非负，且其总和等于 1 的向量称为概率向量。例如，$\boldsymbol{A}=\left[\frac{1}{4},\frac{1}{2},0,\frac{1}{4}\right]$ 就是概率向量。

2) 固定向量

当任一非零向量 $\boldsymbol{u}=(u_1,u_2,\cdots,u_n)$ 左乘某 $n\times n$ 方阵 $\boldsymbol{A}$，其结果仍为 $\boldsymbol{u}$，即

$\boldsymbol{uA}=\boldsymbol{u}$时,$\boldsymbol{u}$为$\boldsymbol{A}$的固定向量。

3) 概率矩阵

由概率向量构成的方阵称为概率矩阵。例如:

$$\boldsymbol{B}=\begin{bmatrix}\frac{1}{2} & \frac{1}{3} & \frac{1}{6}\\ \frac{1}{3} & 0 & \frac{2}{3}\\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4}\end{bmatrix}$$

就是概率矩阵。显然,概率矩阵的元素$0\leqslant b_{ij}\leqslant 1$,且各行之和等于1。

概率矩阵具有以下两个性质:

(1) 若$\boldsymbol{A}$、$\boldsymbol{B}$均为概率矩阵,则$\boldsymbol{AB}$亦为概率矩阵。

(2) 若$\boldsymbol{A}$为概率矩阵,则$\boldsymbol{A}^n$也为概率矩阵。

6.1.4 正规概率矩阵

元素均大于零的概率矩阵称为正规概率矩阵。如

$$\boldsymbol{P}=\begin{bmatrix}\frac{1}{6} & \frac{1}{3} & \frac{1}{2}\\ 0.3 & 0.1 & 0.6\\ 0.01 & 0.90 & 0.09\end{bmatrix}$$

就是一个正规概率矩阵。

正规概率矩阵有如下几个性质:

(1) 正规概率矩阵$\boldsymbol{P}$有一个固定概率向量$\boldsymbol{u}$,且$\boldsymbol{u}$的元素皆为正,此向量称为特征向量。

(2) 正规概率矩阵$\boldsymbol{P}$的各次幂序列$\boldsymbol{P}$, $\boldsymbol{P}^2$, $\boldsymbol{P}^3$, …,将趋近于方阵$\boldsymbol{U}$,且$\boldsymbol{U}$的每一行均为其固定概率向量$\boldsymbol{u}$。

(3) 若$\boldsymbol{F}$为任一概率向量,则向量序列$\boldsymbol{FP}$, $\boldsymbol{FP}^2$, $\boldsymbol{FP}^3$, …,将趋近于$\boldsymbol{P}$的固定概率向量$\boldsymbol{u}$。

6.1.5 转移矩阵

如前所述,事物由状态i经过一步转移到状态j的概率为p_{ij},事物全部一步转移概率的集合所组成的矩阵称为状态转移矩阵。其形式如下:

$$\boldsymbol{P}=\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \tag{6-4}$$

状态转移矩阵必为概率矩阵，其运算规则与概率矩阵相同。如果事物的状态不只经过一步转移，而是经过多步转移，则可用 k 步转移矩阵来表征。记 k 步转移矩阵为 $\boldsymbol{P}^{(k)}$，由定义知：

$$\boldsymbol{P}^{(k)}=\boldsymbol{P}^{k-1}\boldsymbol{P}$$

$$\boldsymbol{P}^{(k)}=\boldsymbol{P}^{k} \tag{6-5}$$

这就说明，k 步转移矩阵只不过是在 $k-1$ 步转移矩阵的基础上再进行一次转移，所以 k 步转移矩阵就是一步转移矩阵的 k 次方。

【例 6-2】 求转移矩阵 $\boldsymbol{P}=\begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix}$ 的二步转移矩阵。

$$\boldsymbol{P}^{(2)}=\boldsymbol{P}^{2}=\begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix}\begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix}=\begin{bmatrix} 0.67 & 0.33 \\ 0.66 & 0.34 \end{bmatrix}$$

6.1.6 正规马尔柯夫链及其稳定状态

若某事物的状态转移矩阵可以表达为正规概率矩阵，则该马尔柯夫链就是正规的，通过若干步转移，最终会达到某种稳定状态，即其后再转移一次、二次、…，结果不再变化，这时稳定状态可用行向量 $\boldsymbol{X}$ 表示，$\boldsymbol{X}=(x_1, x_2, \cdots, x_n)$；$\sum_{i=1}^{n} x_i=1$。

可见，该行向量 $\boldsymbol{X}$ 是此正规概率转移矩阵的固定概率向量。

6.2 马尔柯夫预测模型

6.2.1 建模过程

设事物互不相容的状态有 n 个，事物的初始概率向量 $\boldsymbol{S}^{(0)}$ 为

$$\boldsymbol{S}^{(0)}=[S_1^{(0)}\, S_2^{(0)} \cdots S_n^{(0)}] \tag{6-6}$$

式中，$S_i^{(0)}\,(i=1, 2, \cdots, n)$ 为处于状态 i 的初始概率，且有

$$\sum_{i=1}^{n} S_i^{(0)} = 1$$

若经过 k 步转移后处在状态 i 的概率为 $S_i^{(k)}$，由切普曼-柯尔莫戈洛夫方程可知

$$S_j^{k+1} = \sum_{i=1}^{n} (S_i^{(k)} \cdot p_{ij}) \ (j = 1, 2, \cdots, n) \tag{6-7}$$

写成矩阵形式，即有

$$\begin{aligned}&[S_1^{(k+1)} S_2^{(k+1)} \cdots S_n^{(k+1)}]\\&= [S_1^{(k)} S_2^{(k)} \cdots S_n^{(k)}] \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} (k = 0, 1, 2, \cdots)\end{aligned} \tag{6-8}$$

或

$$\boldsymbol{S}^{k+1} = \boldsymbol{S}^{(k)} \cdot \boldsymbol{P}$$

由递推关系，则有

$$\boldsymbol{S}^{(1)} = \boldsymbol{S}^{(0)} \boldsymbol{P}$$

$$\boldsymbol{S}^{(2)} = \boldsymbol{S}^{(1)} \cdot \boldsymbol{P} = \boldsymbol{S}^{(0)} \cdot \boldsymbol{P}^2$$

$$\cdots\cdots$$

$$\boldsymbol{S}^{(k+1)} = \boldsymbol{S}^{(k)} \cdot \boldsymbol{P} = \boldsymbol{S}^{(0)} \cdot \boldsymbol{P}^{k+1} \tag{6-9}$$

式(6－9)就是$(k+1)$期的马尔柯夫预测模型。所以，式(6－9)又可写成下列形式

$$\boldsymbol{S}^{(k+1)} = \boldsymbol{S}^{(0)} \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}^{(k+1)} \tag{6-10}$$

从式(6－10)可以看出，对于马尔柯夫链，它处于任一时刻的概率可由其初始概率和一步转移概率所决定。

6.2.2 适用条件

马尔柯夫预测模型式(6－9)和式(6－10)，只适用于具有马尔柯夫过程特性的时间序列，并且要求时间序列在预测期间内满足以下条件：

(1) 转移概率矩阵逐期不变，即每一时刻向下一时刻的转移概率都是一样的，

均为一步转移概率。也就是说，马尔柯夫链必是平稳的。

(2) 状态个数保持不变。

(3) 状态的转移只受前一期的影响，而与前一期以前的状态无关。也就是说，该马尔柯夫预测模型只适用于一阶马尔柯夫链的情况。

6.2.3 预测步骤

对符合上述适用条件的预测对象，便能构成一阶马尔柯夫链，并据此建立预测模型。其预测流程如图 6-1 所示。

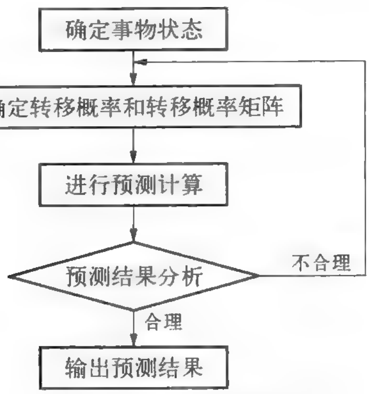


图 6-1 马尔柯夫预测流程

6.3 马尔柯夫预测法的应用

下面用例题说明马尔柯夫预测法的应用。

【例 6-3】 A、B、C、D 货代公司同时在某港口城市从事货运代理业务，由于受服务质量、价格、经营水平及规模等因素影响，每月客户都有变化。经过调查，本年 9 月 1 日各货代公司的客户与 8 月客户情况如表 6-2 所示；8 月 1 日的客户到 9 月 1 日发生变化的情况如表 6-3 所示。要求根据本年 8—9 月的变化，分析预测本年后 3 个月各货代公司客户占有率。

表 6-2 客户情况

货代公司	A	B	C	D
8 月客户	545	495	417	382
9 月 1 日市区客户	65	55	52	54

续 表

货代公司	A	B	C	D
9月1日保留客户 9月1日客户保留率	480 88.1%	440 88.9%	365 87.5%	328 85.9%

表 6-3 客户发生变化情况

货代公司	A	B	C	D	8月客户占有数
A	480	30	20	15	545
B	24	440	16	15	495
C	18	20	365	14	417
D	15	25	14	328	382
9月客户占有数	537	515	415	372	1 839

(1) 确定各货代公司 9 月的客户占有率。

因为

$$\text{客户占有率}=\frac{\text{各货代公司占有客户数}}{\text{总客户数}}$$

所以，A 货代公司客户占有率为

$$\frac{537}{1\,839}=0.292$$

同理可求 B、C、D 货代公司客户占有率分别是 0.280、0.226、0.202。

因此，初始概率向量 $\boldsymbol{S}^{(0)}$ 为

$$\boldsymbol{S}^{(0)}=(0.292\quad 0.280\quad 0.226\quad 0.202)$$

(2) 确定用户转移概率和转移矩阵 $\boldsymbol{P}$。

根据表 6-2 和表 6-3 提供的数据，可计算 8—9 月客户转移概率。其中，A 货代公司客户转移概率为

$$p_{11}=\frac{480}{545}=0.881,\ p_{12}=\frac{30}{545}=0.055$$

$$p_{13}=\frac{20}{545}=0.037,\ p_{14}=\frac{15}{545}=0.027$$

B 货代公司客户转移概率为

$$p_{21}=\frac{24}{495}=0.049,\ p_{22}=\frac{440}{495}=0.889$$

$$p_{23}=\frac{16}{495}=0.032,\ p_{24}=\frac{15}{495}=0.030$$

同样可计算出C、D货代公司的客户转移概率。将上述计算结果写成矩阵就是8—9月的一步转移概率矩阵

$$\boldsymbol{P}=\begin{bmatrix}0.881 & 0.055 & 0.037 & 0.027\\ 0.049 & 0.889 & 0.032 & 0.030\\ 0.043 & 0.048 & 0.875 & 0.034\\ 0.039 & 0.065 & 0.037 & 0.859\end{bmatrix}$$

(3) 进行预测计算。

若本年后3个月各月之间客户转移概率不变,则可运用式(6-9)进行预测。10月各货代公司的客户占有率为

$$\boldsymbol{S}^{(1)}=\boldsymbol{S}^{(0)}\boldsymbol{P}=[0.292\quad 0.280\quad 0.226\quad 0.202]\begin{bmatrix}0.881 & 0.055 & 0.037 & 0.027\\ 0.049 & 0.889 & 0.032 & 0.030\\ 0.043 & 0.048 & 0.875 & 0.034\\ 0.039 & 0.065 & 0.037 & 0.859\end{bmatrix}$$

$$=[0.289\quad 0.289\quad 0.225\quad 0.197]$$

11月客户占有率为

$$\boldsymbol{S}^{(2)}=\boldsymbol{S}^{(1)}\boldsymbol{P}=[0.289\quad 0.289\quad 0.225\quad 0.197]\begin{bmatrix}0.881 & 0.055 & 0.037 & 0.027\\ 0.049 & 0.889 & 0.032 & 0.030\\ 0.043 & 0.048 & 0.875 & 0.034\\ 0.039 & 0.065 & 0.037 & 0.859\end{bmatrix}$$

$$=[0.286\quad 0.297\quad 0.224\quad 0.193]$$

12月客户占有率为

$$\begin{aligned}\boldsymbol{S}^{(3)}&=\boldsymbol{S}^{(2)}\boldsymbol{P}=[0.286\quad 0.297\quad 0.224\quad 0.193]\boldsymbol{P}\\ &=[0.284\quad 0.303\quad 0.223\quad 0.190]\end{aligned}$$

预测结果表明,如果各货代公司占有客户的变化仍按本年8—9月的规律进行,则到本年年底,原来客户占有率比较接近的4个货代公司将产生较大差异。B货代公司的客户占有率将明显高于其他货代公司,由9月的第二位跃居第一位,而D货代公司的客户占有率则明显低于其他货代公司。

由上述预测结果可知，按目前这种转移概率矩阵，对D货代公司不利。D货代公司要想改变这种状况，必须设法改变客户转移概率矩阵。决定转移概率的因素主要有提高服务质量、降低服务价格、改善经营手段及加强宣传等。假如D货代公司从9月采取低价策略，因此，转移概率矩阵根据原有数据资料要重新估计。如果变化为

$$\begin{bmatrix} 0.881 & 0.055 & 0.037 & 0.027 \\ (-0.010) & & & (+0.010) \\ 0.049 & 0.889 & 0.032 & 0.030 \\ & (-0.008) & & (+0.008) \\ 0.043 & 0.048 & 0.875 & 0.034 \\ & & (-0.013) & (+0.013) \\ 0.039 & 0.065 & 0.037 & 0.859 \\ (-0.010) & (-0.015) & (-0.015) & (+0.04) \end{bmatrix}$$

这就是说，D货代公司9月降价后，从A货代公司争得1%的客户；从B货代公司争得0.8%的客户；从C货代公司争得1.3%的客户。同时，D货代公司客户保留率提高4%，而转向A货代公司的客户减少1%，转向B、C货代公司的客户均减少1.5%，这样一来，转移概率矩阵变为

$$\boldsymbol{P}=\begin{bmatrix} 0.871 & 0.055 & 0.037 & 0.037 \\ 0.049 & 0.881 & 0.032 & 0.038 \\ 0.043 & 0.048 & 0.862 & 0.047 \\ 0.029 & 0.050 & 0.022 & 0.899 \end{bmatrix}$$

这时，10月的各货代公司客户占有率为

$$\begin{aligned} \boldsymbol{S}^{(1)} &= \boldsymbol{S}^{(0)}\boldsymbol{P}=[0.292 \quad 0.280 \quad 0.226 \quad 0.202]\boldsymbol{P} \\ &= [0.283 \quad 0.284 \quad 0.219 \quad 0.214] \end{aligned}$$

原D货代公司10月客户占有率为19.7%，即在1 839个客户中只占有362户，降价后10月客户为394户，说明采取降价策略后，D货代公司客户增加32户。如果降价代价为10万元，而每户平均销售获利0.5万元，则10月D货代公司将净获利0.5×32−10=6(万元)。

按此规律变化，11月、12月各货代公司的客户占有率分别为

$$\begin{aligned} \boldsymbol{S}^{(2)} &= \boldsymbol{S}^{(1)}\boldsymbol{P}=[0.283 \quad 0.284 \quad 0.219 \quad 0.214]\boldsymbol{P} \\ &= [0.276 \quad 0.287 \quad 0.213 \quad 0.224] \end{aligned}$$

$$\begin{aligned}\boldsymbol{S}^{(3)} &= \boldsymbol{S}^{(2)}\boldsymbol{P} = [0.276 \quad 0.287 \quad 0.213 \quad 0.224]\boldsymbol{P} \\ &= [0.270 \quad 0.289 \quad 0.208 \quad 0.233]\end{aligned}$$

预测结果表明，若D货代公司9月采取降价策略，则本年后3个月，D货代公司的客户占有率将会大有改观。

【例6-4】 某集装箱码头A装卸设备的状态转移概率如表6-4所示。试预测该设备一天后、两天后、三天后和四天后的状态概率。

表6-4 状态转移概率

从＼至	无故障(状态①)	有故障(状态②)
无故障(状态①)	0.7	0.3
有故障(状态②)	0.6	0.4

假设A设备开始处于状态①，则初始状态概率为 $S^{(0)} = [1 \quad 0]$。

因转移矩阵 $\boldsymbol{P} = \begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix}$，则：

一天后的状态概率为

$$\boldsymbol{S}^{(1)} = \boldsymbol{S}^{(0)}\boldsymbol{P} = [1 \quad 0]\begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix} = [0.7 \quad 0.3]$$

两天后的状态概率为

$$\boldsymbol{S}^{(2)} = \boldsymbol{S}^{(0)}\boldsymbol{P}^2 = \boldsymbol{S}^{(1)}\boldsymbol{P} = [0.7 \quad 0.3]\begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix} = [0.67 \quad 0.33]$$

三天后的状态概率为

$$\boldsymbol{S}^{(3)} = \boldsymbol{S}^{(0)}\boldsymbol{P}^3 = \boldsymbol{S}^{(2)}\boldsymbol{P} = [0.67 \quad 0.33]\begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix} = [0.667 \quad 0.333]$$

四天后的状态概率为

$$\boldsymbol{S}^{(4)} = \boldsymbol{S}^{(0)}\boldsymbol{P}^3 = \boldsymbol{S}^{(3)}\boldsymbol{P} = [0.667 \quad 0.333]\begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix} = [0.6667 \quad 0.3333]$$

如果A设备开始处于状态②，则初始状态概率为 $\boldsymbol{S}^{(0)} = [0 \quad 1]$。

因转移矩阵 $\boldsymbol{P} = \begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix}$，则一天后的状态概率为

$$S^{(1)}=S^{(0)}P=[0\quad 1]\begin{bmatrix}0.7 & 0.3\\ 0.6 & 0.4\end{bmatrix}=[0.6\quad 0.4]$$

两天后的状态概率为

$$S^{(2)}=S^{(0)}P^2=S^{(1)}P=[0.6\quad 0.4]\begin{bmatrix}0.7 & 0.3\\ 0.6 & 0.4\end{bmatrix}=[0.66\quad 0.34]$$

三天后的状态概率为

$$S^{(3)}=S^{(0)}P^3=S^{(2)}P=[0.66\quad 0.34]\begin{bmatrix}0.7 & 0.3\\ 0.6 & 0.4\end{bmatrix}=[0.666\quad 0.334]$$

四天后的状态概率为

$$S^{(4)}=S^{(0)}P^4=S^{(3)}P=[0.666\quad 0.334]\begin{bmatrix}0.7 & 0.3\\ 0.6 & 0.4\end{bmatrix}=[0.6666\quad 0.3334]$$

由此例可以看出，无论 A 设备开始处于什么状态，该设备未来状态①的概率将趋近于 2/3，而处于状态②的概率将趋近于 1/3。这两个概率分别为 A 设备在状态①和在状态②的稳定状态概率。下面将介绍马尔柯夫链的稳定状态概率解法及应用。

6.4 马尔柯夫链的稳态概率及应用

马尔柯夫链在一定条件下，经过 k 步(k 足够大)转移后，会达到稳定状态，且与初始状态无关。马尔柯夫链达到稳定状态时的状态概率就是稳定状态概率，简称稳态概率；又因稳定状态与初始状态无关，满足遍历性，因而也称遍历概率。在预测中，只要求得稳态概率，就能以此进行预测。

6.4.1 马尔柯夫链存在稳定状态的条件

如前所述，若概率矩阵中，诸元素均为正值，则该概率矩阵为正规概率矩阵。即对于矩阵 $\boldsymbol{A}$，若 $\boldsymbol{A}=(a_{ij})_{n\times n}$ 满足

$$a_{ij}>0 \text{ 且} \sum_{j=1}^{n}a_{ij}=1\ (i,\ j=1,\ 2,\ \cdots,\ n) \tag{6-11}$$

则矩阵 $\boldsymbol{A}$ 为正规概率矩阵。可以证明，若马尔柯夫链的一步状态转移矩阵为正规概率矩阵，那么马尔柯夫链必定有稳定状态。通常情况下，市场占有率、设备状态、期望利润等的状态转移概率矩阵恰好满足正规概率矩阵的条件。这时的关键在于如何求解马尔柯夫链稳态概率。

6.4.2 马尔柯夫链稳态概率解法

由马尔柯夫链稳定状态的定义可知，当其处于稳定状态时，有 $\boldsymbol{S}^{(k+1)} = \boldsymbol{S}^{(k)}$，且 $\sum_{i=1}^{n} S_i^{(k)} = 1$，则

$$\boldsymbol{S}^{(k+1)} = \boldsymbol{S}^{(k)}\boldsymbol{P} = \boldsymbol{S}^{(k)} \tag{6-12}$$

将式(6-12)写成矩阵形式为

$$\begin{aligned}&[S_1^{(k)} \quad S_2^{(k)} \quad \cdots \quad S_n^{(k)}]\\ &= [S_1^{(k)} \quad S_2^{(k)} \quad \cdots \quad S_n^{(k)}]\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}\end{aligned} \tag{6-13}$$

将式(6-13)展开，并加上 $S_1^{(k)} + S_2^{(k)} + \cdots + S_n^{(k)} = 1$，得

$$\begin{cases} S_1^{(k)} = S_1^{(k)} p_{11} + S_2^{(k)} p_{21} + \cdots + S_n^{(k)} p_{n1} \\ S_2^{(k)} = S_1^{(k)} p_{12} + S_2^{(k)} p_{22} + \cdots + S_n^{(k)} p_{n2} \\ \qquad\cdots\cdots \\ S_n^{(k)} = S_1^{(k)} p_{1n} + S_2^{(k)} p_{2n} + \cdots + S_n^{(k)} p_{nn} \\ S_1^{(k)} + S_2^{(k)} + \cdots + S_n^{(k)} = 1 \end{cases} \tag{6-14}$$

式中，有 n 个变量，但有 $n+1$ 个方程，说明其中有一个方程式多余的，从而可删去其中任一个方程(不能删去 $S_1^{(k)} + S_2^{(k)} + \cdots + S_n^{(k)} = 1$)，若删去第 $n-1$ 个方程，移项整理得到

$$\begin{cases} (p_{11} - 1)S_1^{(k)} + p_{21}S_2^{(k)} + \cdots + p_{n1}S_n^{(k)} = 0 \\ p_{12}S_1^{(k)} + (p_{22} - 1)S_2^{(k)} + \cdots + p_{n2}S_n^{(k)} = 0 \\ \qquad\cdots\cdots \\ p_{1n}S_1^{(k)} + p_{2n}S_2^{(k)} + \cdots + (p_{nn} - 1)S_n^{(k)} = 0 \\ S_1^{(k)} + S_2^{(k)} + \cdots + S_n^{(k)} = 1 \end{cases} \tag{6-15}$$

再将式(6-15)写成矩阵形式，得

$$\begin{bmatrix} p_{11}-1 & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22}-1 & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn}-1 \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} S_1^{(k)} \\ S_2^{(k)} \\ \vdots \\ S_n^{(k)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \tag{6-16}$$

设

$$\boldsymbol{P}_1 = \begin{bmatrix} p_{11}-1 & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22}-1 & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn}-1 \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

$$\boldsymbol{S}^{(k)} = \begin{bmatrix} S_1^{(k)} \\ S_2^{(k)} \\ \vdots \\ S_n^{(k)} \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

则有

$$\boldsymbol{P}_1 \boldsymbol{S}^{(k)} = \boldsymbol{B}$$

因此有

$$\boldsymbol{S}^{(k)} = \boldsymbol{P}_1^{-1} \boldsymbol{B}$$

式(6－17)求得的 $\boldsymbol{S}^{(k)}$ 就是马尔柯夫链的稳态矩阵。

【例 6－5】 试计算【例 6－4】中 A 装卸设备的稳态矩阵。

【例 6－4】中，$\boldsymbol{P} = \begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix}$，从而得

$$\boldsymbol{P}_1 = \begin{bmatrix} -0.3 & 0.6 \\ 1 & 1 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

所以

$$\boldsymbol{S}^{(k)} = \begin{bmatrix} -0.3 & 0.6 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{1}{-0.9} \begin{bmatrix} 1 & -0.6 \\ 1 & -0.3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{1}{-0.9} \begin{bmatrix} -0.6 \\ -0.3 \end{bmatrix} = \begin{bmatrix} 2/3 \\ 1/3 \end{bmatrix}$$

即
$$S_1^{(k)} = 2/3,\ S_2^{(k)} = 1/3$$
因此，A 装卸设备处于状态①的稳态概率为 2/3，处于状态②的稳态概率为 1/3。

思考与练习

1. 有 3 家企业 A、B、C，由于产品质量、服务质量、价格、促销、分销等原因，订购用户的变化如下：

4 月：A 家 200 户，B 家 500 户，C 家 300 户。

5 月：A 家保留 160 户，而从 B 转入 35 户，从 C 转入 25 户；B 家保留 450 户，而从 A 转入 20 户，从 C 转入 20 户；C 家保留 255 户，从 A 转入 20 户，从 B 转入 15 户。

试求其转移概率矩阵。

2. 某产品每季度的市场销售状态为畅销、滞销两种。6 年来 24 个季度的状态如题表 6－1 所示。

题表 6－1

季度	1	2	3	4	5	6	7	8	9	10	11	12
状态	畅	畅	滞	畅	滞	滞	畅	畅	畅	滞	畅	滞
季度	13	14	15	16	17	18	19	20	21	22	23	24
状态	畅	畅	滞	滞	畅	畅	滞	畅	滞	畅	畅	畅

试求市场状态的一步和二步转移矩阵 $\boldsymbol{P} = \begin{bmatrix} 0.5 & 0.5 \\ 0.4 & 0.6 \end{bmatrix}$。

7 灰色预测法

预测就是借助于对过去的探讨去推测、了解未来。灰色预测通过原始数据的处理和灰色模型的建立，发现、掌握系统发展规律，对系统的未来状态做出科学的定量预测。对于一个具体的问题，究竟选择什么样的预测模型应以充分的定性分析结论为依据。模型的选择不是一成不变的。一个模型要经过多种检验才能判定其是否合适。只有通过检验的模型才能用来进行预测。本章将简要介绍灰数、灰色预测的概念，灰色预测模型的构造、检验、应用，最后阐述灾变预测的原理。

7.1 灰数简介

7.1.1 灰数

灰色系统理论中的一个重要概念是灰数。灰数是指未明确指定的数，即处在某一范围内的数，灰数是区间数的一种推广。

灰色系统用灰数、灰色方程、灰色矩阵等来描述，其中灰数是灰色系统的基本“单元”或“细胞”。

我们把只知道大概范围而不知其确切值的数称为灰数。在应用中，灰数实际上指在某一个区间或某个一般的数集内取值的不确定数，通常用记号“$\otimes$”表示灰数。

灰数有以下几类：

1）仅有下界的灰数

有下界而无上界的灰数记为 $\otimes\in[\underline{a},\infty]$ 或 $\otimes(\underline{a})$。其中$\underline{a}$ 为灰数$\otimes$的下确界，它是一个确定的数，我们称$[\underline{a},\infty]$为$\otimes$的取数域，简称$\otimes$的灰域。

一棵生长着的大树，其质量便是有下界的灰数，因为大树的质量必大于零，但不可能用一般手段知道其准确的质量，若用$\otimes$表示大树的质量，便有 $\otimes=[0,\infty]$。

2）仅有上界的灰数

有上界而无下界的灰数记为 $\otimes\in(-\infty,\bar{a}]$ 或 $\otimes(\bar{a})$。其中 $\bar{a}$ 为灰数$\otimes$的上确界，是一个确定的数。

一项港口投资工程，要有个最高投资限额，一台港口设备要有个承受电压或通过电流的最高临界值。工程投资、电器设备的电压、电流容许值都是有上界的灰数。

3）区间灰数

既有下界$\underline{a}$又有上界$\bar{a}$的灰数称为区间灰数，记为$\otimes \in [\underline{a}, \bar{a}]$。

海豹的质量在20～25 kg之间，某人的身高在1.8～1.9 m之间，可分别记为

$$\otimes_1 \in [20, 25], \otimes_2 \in [1.8, 1.9]$$

4）连续灰数与离散灰数

在某一区间内取有限个值或可数个值的灰数称为离散灰数，取值连续地充满某一区间的灰数称为连续灰数。

某人的年龄在30～35岁之间，此人的年龄可能是30,31,32,33,34,35这几个数，因此年龄是离散灰数。人的身高、体重等是连续灰数。

5）黑数与白数

当$\otimes \in [-\infty, \infty]$或$\otimes \in [\otimes_1, \otimes_2]$，即当$\otimes$的上、下界皆为无穷或上、下界都是灰数时，称$\otimes$为黑数。

当$\otimes \in [\underline{a}, \bar{a}]$且$\underline{a} = \bar{a}$时，称$\otimes$为白数。

为方便讨论，我们将黑数与白数看成特殊的灰数。

6）本征灰数与非本征灰数

本征灰数是指不能或暂时还不能找到一个白数作为其“代表”的灰数，比如一般的事前预测值、宇宙的总能量、准确到秒或微妙的“年龄”等都是本征灰数。

非本征灰数是指凭先验信息或某种手段，可以找到一个白数作为其“代表”的灰数。我们称此白数为相应灰数的白化值，记为$\tilde{\otimes}$，并用$\otimes(a)$表示以a为白化值的灰数。如扩建码头泊位预算1 000万元，可将1 000万元作为泊位扩建预算$\otimes(1\,000)$的白化数，记为$\tilde{\otimes}(1\,000) = 1\,000$。

从本质上来看，灰数又可分为信息型、概念型、层次型三类：

(1) 信息型灰数，指因暂时缺乏信息而不能肯定其取值的数，如：预计某港口今年总吞吐量在100万吨以上，$\otimes \in [100, \infty]$；估计某船公司今年利润总额将达(7 000～9 000)万元，$\otimes \in [7\,000, 9\,000]$；预计西安地区5月最高气温不超过36℃，$\otimes \in [0, 36]$。这些都是信息型灰数。由于暂时缺乏信息，不能肯定某数的确切取值，而到一定的时间，通过信息补充，灰数可以完全变白。

(2) 概念型灰数，也称意愿型灰数，是指由人们的某种观念、意愿形成的灰数。如某人希望至少获得1万元科研经费，并且越多越好，$\otimes \in [10\,000, \infty]$；某集装箱码头承诺外集卡在港等待时间为1小时，希望大幅度降低，当然越小越好，$\otimes \in$

[0, 1]。这些都是概念型灰数。

(3) 层次型灰数,由层次的改变形成的灰数。有的数,从系统的高层次,即宏观层次、整体层次或认识的概括层次上看是白的,可到低层次上,即到系统的微观层次、分部层次或认识的深化层次则可能是灰的。例如,一个集装箱码头的堆场面积,以平方米度量是白的,若精确到万分之一平方毫米就成灰的了。

7.1.2 灰数白化与灰度

有一类灰数是在某个基本值附近变动的,这类灰数白化比较容易,我们可以其基本值为主要白化值。以 a 为基本值的灰数可记为 $\otimes(a)=a+\delta_a$ 或 $\otimes(a)\in(-\infty, a, +\infty)$,其中 δ_a 为扰动灰元,此灰数的白化值为 $\tilde{\otimes}(a)=a$。如今年的科研经费在 5 万元左右,可表示为 $\otimes(50\,000)=50\,000+\delta$,或 $\otimes(50\,000)\in(-\infty, 50\,000, +\infty)$,它的白化值为 50 000。

对于一般的区间灰数 $\otimes\in[a, b]$,我们将白化值 $\tilde{\otimes}$ 取为

$$\tilde{\otimes}=\alpha a+(1-\alpha)b,\ \alpha\in[0, 1]$$

形如 $\tilde{\otimes}=\alpha a+(1-\alpha)b,\ \alpha\in[0, 1]$ 的白化称为等权白化。

在等权白化中,取 $\alpha=\dfrac{1}{2}$ 而得到的白化值称为等权均值白化。

当区间灰数取值的分布信息缺乏时,常采用等权均值白化。设区间灰数

$$\otimes_1\in[a, b],\ \tilde{\otimes}_1=\alpha a+(1-\alpha)b,\ \alpha\in[0, 1]$$
$$\otimes_2\in[c, d],\ \tilde{\otimes}_2=\beta c+(1-\beta)d,\ \beta\in[0, 1]$$

当 $\alpha=\beta$ 时,称 $\otimes_1$ 与 $\otimes_2$ 取数一致;当 $\alpha\neq\beta$ 时,称 $\otimes_1$ 与 $\otimes_2$ 取数非一致。

在灰数的分布信息已知时,往往采取非等权白化。例如,某人 2000 年的年龄可能是 40~60 岁,$\otimes\in[40, 60]$ 是个灰数。根据了解,此人受初、中级教育共 12 年,并且是在 60 年代中期考入大学的,故此人的年龄到 2000 年为 58 岁左右的可能性较大,或者说在 56~60 岁的可能性较大。这样的灰数,如果再做等权白化,显然是不合理的。为此,我们用白化权函数来描述一个灰数对其取值范围内不同数值的"偏爱"程度。

对概念型灰数中表示意愿的灰数,其白化权函数一般设计为单调增函数。

一般来说,一个灰数的白化权函数是研究者根据已知信息设计的,没有固定的程式。函数曲线的起点和终点一般应有其含义。如在外贸谈判中,就有一个由灰变白的过程。开始谈判时,甲方说我的出口额至少要 5 亿元,乙方说我的进口额不大于 3 亿元。则成交额这一灰数将在 3 亿元~5 亿元间取值,其白化权函数可将起

点定为 3 亿元，终点定为 5 亿元。

灰度即为灰数的测度。灰数的灰度在一定程度上反映了人们对灰色系统之行为特征的未知程度。在实际应用中，我们会遇到大量的白化权函数未知的灰数，例如由一般灰色系统之行为特征预测值构成的灰数，就难以给出其白化权函数。

我们认为，灰数的灰度主要与相应定义信息域的长度及其基本值有关。如果考虑一个 4 000 左右的灰数，给出其估计值的两个灰数 $\otimes_1 \in [3\,998,\ 4\,002]$ 和 $\otimes_2 \in [3\,900,\ 4\,100]$，显然$\otimes_1$ 比$\otimes_2$ 更有价值，亦即$\otimes_1$ 比$\otimes_2$ 灰度小，若再考虑一个基本值为 4 的灰数，给出灰数 $\otimes_3 \in [2,\ 6]$，虽然$\otimes_1$ 与$\otimes_3$ 的长度都是 4，但$\otimes_1$ 比$\otimes_3$ 的灰度小是显而易见的。

7.2 灰色预测的概念

7.2.1 灰色系统及灰色预测的概念

7.2.1.1 灰色系统基本概念

灰色系统产生于控制理论的研究中。

若一个系统的内部特征是完全已知的，即系统的信息是充足完全的，我们称之为白色系统。

若一个系统的内部信息是一无所知，一团漆黑，只能从它同外部的联系来观测研究，这种系统便是黑色系统。

灰色系统介于两者之间，灰色系统的一部分信息是已知的，一部分是未知的。

区别白色和灰色系统的重要标志是系统各因素间是否有确定的关系。

在工程技术、经济等各种系统中经常会遇到信息不完全的情况。比如：一项土建工程，尽管材料、设备、施工计划、图纸是齐备的，可是还很难估计施工进度与质量，这是缺乏劳动力及技术水平的信息；一般经济系统，除了输出的时间数据列（比如产值、产量、总收入、总支出等）外，其输入数据列不明确或者缺乏，因而难以建立确定的完整的模型，这是缺乏系统信息；工程系统是客观实体，有明确的“内”、“外”关系（即系统内部与系统外部，或系统本体与系统环境），可以较清楚地明确输入与输出，因此可以较方便地分析输入对输出的影响，可是港口经济系统是抽象的对象，没有明确的“内”、“外”关系，不是客观实体，因此就难以分析输入（投入）对输出（产出）的影响，这是缺乏“模型信息”（即用什么模型，用什么量进行观测控制等信息）。信息不完全的情况归纳起来有：元素（参数）信息不完全；结构信息不完全；关系信息（特指“内”、“外”关系）不完全；运行的行为信息不完全。

例如，一个小型货运代理企业可看作是一个系统，在人员、资金、损耗、业绩信

息完全明确的情况下，可算出该企业的盈利大小多少，可以判断未来的发展态势、资金的周转速度等，这样的系统是白色系统。

遥远的某个星球，也可以看作一个系统，虽然知道其存在，但体积多大、质量多少、距离地球多远，这些信息完全不知道，这样的系统是黑色系统。

人体是一个系统，人体的一些外部参数（如身高、体温、脉搏等）是已知的，而其他一些参数，如人体的穴位有多少，穴位的生物、化学、物理性能，生物的信息传递等尚未知道透彻，这样的系统是灰色系统。

显然，黑色、灰色、白色都是一种相对的概念。世界上没有绝对的白色系统，因为任何系统总有未确知的部分，也没有绝对的黑色系统，因为既然一无所知，也就无所谓该系统的存在了。

7.2.1.2　*灰色系统的特点*

灰色系统理论以“部分信息已知、部分信息未知”的“小样本”、“贫信息”不确定型系统为研究对象。

1）用灰色数学来处理不确定量，使之量化

在数学发展史上，最早研究的是确定型的微分方程，即在拉普拉斯决定论框架内的数学。他认为一旦有了描写事物的微分方程及初值，就能确知事物任何时候的运动。随后发展了概率论与数理统计，用随机变量和随机过程来研究事物的状态和运动。模糊数学则研究没有清晰界限的事物，如枢纽港和支线港之间没有确定的吞吐量界限加以截然划分等，它通过隶属函数来使模糊概念量化，因此能用模糊数学来描述如语言、不精确推理以及若干人文科学。灰色系统理论则认为不确定量是灰数，用灰色数学来处理不确定量，同样能使不确定量予以量化。如图 7－1 所示。

$$\text{不确定量}\xrightarrow{1,\ 2,\ 3}\text{量化（用确定量的方法研究）}$$

图 7－1　处理不确定量的方法

1—概率论与数理统计；2—模糊数学；3—灰色数学（灰色系统理论）

2）充分利用已知信息寻求系统的运动规律

研究灰色系统的关键是如何使灰色系统白化、模型化、优化。

灰色系统视不确定量为灰色量。提出了灰色系统建模的具体数学方法，它能利用时间序列来确定微分方程的参数。灰色预测不是把观测到的数据序列视为一个随机过程，而是看作随时间变化的灰色量或灰色过程，通过累加生成和累减生成逐步使灰色量白化，从而建立相应于微分方程解的模型并做出预测。这样，对某些大系统和长期预测问题，就可以发挥作用。

3) 灰色系统理论能处理贫信息系统

灰色预测模型只要求较短的观测资料即可，这和时间序列分析，多元分析等概率统计模型要求较长资料很不一样。因此，对于某些只有少量观测数据的项目来说，灰色预测是一种有用的工具。

7.2.1.3　灰色预测

灰色系统分析方法是通过鉴别系统因素之间发展趋势的相似或相异程度，即进行关联度分析，并通过对原始数据的生成处理来寻求系统变动的规律。生成数据序列有较强的规律性，可以用它来建立相应的微分方程模型，从而预测事物未来的发展趋势和未来状态。

灰色预测是用灰色模型 GM(1, 1)来进行定量分析的，通常分为以下几类：

(1) 灰色时间序列预测。用等时距观测到的反映预测对象特征的一系列数量(如产量、销量、吞吐量、人口数量、出口额、利率等)构造灰色预测模型，预测未来某一时刻的特征量，或者达到某特征量的时间。

(2) 畸变预测(灾变预测)。通过模型预测异常值出现的时刻，预测异常值什么时候出现在特定时区内。

(3) 波形预测，或称为拓扑预测，它是通过灰色模型预测事物未来变动的轨迹。

(4) 系统预测，是对系统行为特征指标建立一族相互关联的灰色预测理论模型，在预测系统整体变化的同时，预测系统各个环节的变化。

上述灰预测方法的共同特点是：

(1) 允许少数据预测。

(2) 允许对灰因果律事件进行预测。例如：

■ 灰因白果律事件。在集装箱码头吞吐预测中，影响吞吐量大小的因子很多，多到无法枚举，故为灰因，然而吞吐量却是具体的，故为白果。吞吐量预测即为灰因白果律事件预测。

■ 白因灰果律事件。在开发项目前景预测时，开发项目的投入是具体的，为白因，而项目的效益暂时不很清楚，为灰果。项目前景预测即为白因灰果律事件预测。

(3) 具有可检验性，包括：建模可行性的级比检验(事前检验)，建模精度检验(模型检验)，预测的滚动检验(预测检验)。

7.2.2　预备知识

7.2.2.1　生成数

生成数分为累加生成数(1－AGO)与累减生成数(IAGO)。

(1) 累加生成数,1－AGO 指一次累加生成。

记原始序列为

$$X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)\}$$

生成序列为

$$X^{(1)} = x\{x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n)\}$$

上标“0”表示原始序列,上标“1”表示一次累加生成序列。其中:

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i) = x^{(1)}(k-1) + x^{(0)}(k)$$

(2) 累减生成数(IAGO),是累加生成的逆运算。

记原始序列为 $X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n)\}$,对 $X^{(1)}$ 做一次累减生成,则得生成序列 $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)\}$,其中,$x^{(0)}(k) = x^{(1)}(k) - x^{(1)}(k-1)$,规定 $x^{(1)}(0) = 0$。

累加生成与累减生成之间的关系如图 7－2 所示。

$$X^{(0)} \xrightarrow{1-\mathrm{AGO}} X^{(1)} \xrightarrow{\mathrm{IAGO}} X^{(0)}$$

图 7－2　累加生成与累减生成之间的关系

7.2.2.2　关联度

为了定量地研究两个事物间的关联程度,人们提出了各种形式的指数,如相关系数和相似系数等。这些指数大多以数理统计原理为基础,需要足够的样本个数或者要求数据服从一定的概率分布。

在客观世界中,有许多因素之间的关系是灰色的,分不清哪些因素之间关系密切,哪些不密切,这样就难以找到主要矛盾和主要特性。灰因素关联分析,目的是定量地表征诸因素之间的关联程度,从而揭示灰色系统的主要特性。关联分析是灰色系统分析和预测的基础。

关联分析是一种相对性的排序分析。从思路上来看,源于几何直观。如图 7－3 所示的 A、B、C、D 四个时间序列,曲线 A 与 B 比较平行,我们就认为 A 与 B 的关联程度大。曲线 C 与 A 随时间变化的方向很不一致,认为 A 与 C 的关联程度较小。曲线 A 与 D 相差最大,则认为两者的关联程度最小。

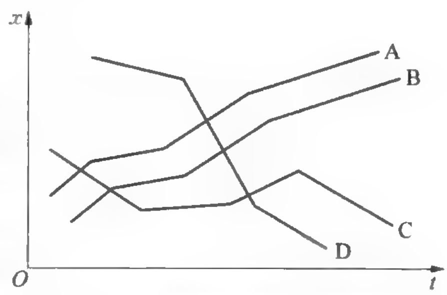


图 7－3　时间序列的几何关联性

将曲线 A 与 B、C、D 的关联程度分别记为 r_{AB}、r_{AC}、r_{AD}，则它们之间有如下排序关系：$r_{AB} > r_{AC} > r_{AD}$，相应的序列 $\{r_{AB}, r_{AC}, r_{AD}\}$ 称为关联序。

由此可见，关联分析实质上是一种曲线间几何形状的分析比较，即几何形状越接近，则发展变化趋势越接近，关联程度越大；反之亦然。

关联度分析是分析系统中各因素关联程度的方法。计算关联度需先计算关联系数。

1) 关联系数的计算

设参考序列为

$$X_0 = \{x_0(1), x_0(2), \cdots, x_0(n)\}$$

比较序列为

$$X_i = \{x_i(1), x_i(2), \cdots, x_i(n)\}$$

关联系数定义为

$$\eta_i(k) = \frac{\min\limits_j \min\limits_l \mid x_0(l) - x_j(l) \mid + P \max\limits_j \max\limits_l \mid x_0(l) - x_j(l) \mid}{\mid x_0(k) - x_i(k) \mid + P \max\limits_j \max\limits_l \mid x_0(l) - x_j(l) \mid} \tag{7-1}$$

式中，$|x_0(k) - x_i(k)|$ 为第 k 点 x_0 与 x_i 的绝对差；$\min\limits_j \min\limits_l \mid x_0(l) - x_j(l) \mid$ 为两级最小差，其中 $\min\limits_l \mid x_0(l) - x_j(l) \mid$ 为第一级最小差，表示在 X_j 序列上找各点与 X_0 最小差；$\min\limits_j \min\limits_l \mid x_0(l) - x_j(l) \mid$ 为第二级最小差，表示在各序列中找出的最小差基础上寻求所有序列中的最小差；$\max\limits_j \max\limits_l \mid x_0(l) - x_j(l) \mid$ 为两级最大差，其含义与最小差相似；P 为分辨率，$0 < P < 1$，一般采用 $P = 0.5$。

对单位不一、初值不同的序列，在计算关联系数之前应首先进行初值化，即将该序列的所有数据分别除以第一数据，将变量化为无单位的相对数值。

2) 关联度的计算

关联系数只表示了各个时刻参考序列和比较序列之间的关联程度，为了从总体上了解序列之间的关联程度，必须求出它们的时间平均值，即关联度。

因此，计算关联度的公式为

$$r_i = \frac{1}{n} \sum_{k=1}^{n} \eta_i(k) \tag{7-2}$$

另外，定量地表征灰色系统诸因子之间关联程度的指数有两种，按其计算方法的差异，分别称为绝对值关联度和速率关联度。以上我们所介绍的是绝对值关联度的概念和计算，有关速率关联度问题，在此不做详述。

7.3 灰色预测模型

7.3.1 GM(1, 1)模型

令 $X^{(0)}$ 为 GM(1, 1)建模序列:

$$X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)\}$$

$X^{(1)}$ 为 $X^{(0)}$ 的 1—AGO 序列:

$$X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n)\}$$

$$X^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i),\ k = 1, 2, \cdots, n$$

令 $Z^{(1)}$ 为 $X^{(1)}$ 的紧邻均值(MEAN)成长序列:

$$Z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \cdots, x^{(1)}(n))$$

$$z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1)$$

则 GM(1, 1)的定义型,即 GM(1, 1)灰微分方程模型为

$$x^{(0)}(k) + az^{(1)}(k) = b \tag{7-3}$$

式中,a 为发展系数;b 为灰色作用量。

模型符号含义如图 7-4 所示。

G	M	(1,	1)
↑	↑	↑	↑
Grey	Model	1 阶方程	1 个变量

图 7-4 GM(1, 1)模型符号含义

设 $\hat{a}$ 为待估参数向量,即 $\hat{a} = (a, b)^{\mathrm{T}}$,则灰色微分方程(7-3)的最小二乘估计参数列满足:

$$\hat{\boldsymbol{a}} = (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{Y}_n$$

式中,

$$\boldsymbol{B} = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix},\ \boldsymbol{Y}_n = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}$$

称

$$\frac{\mathrm{d}x^{(1)}}{\mathrm{d}t}+ax^{(1)}=b \tag{7-4}$$

为灰色微分方程 $x^{(0)}(k)+az^{(1)}(k)=b$ 的白化方程,也称为影子方程。如上所述,则有

(1) 白化方程 $\frac{\mathrm{d}x^{(1)}}{\mathrm{d}t}+ax^{(1)}=b$ 的解也称时间响应函数为

$$\hat{x}^{(1)}(t)=\left(x^{(1)}(0)-\frac{b}{a}\right)\mathrm{e}^{-at}+\frac{b}{a}$$

(2) GM(1, 1)灰色微分方程 $x^{(0)}(k)+az^{(1)}(k)=b$ 的时间响应序列为

$$\hat{x}(k+1)=\left[x^{(1)}(0)-\frac{b}{a}\right]\mathrm{e}^{-ak}+\frac{b}{a}\quad(k=1,\ 2,\ \cdots,\ n)$$

(3) 取 $x^{(0)}(0)=x^{(0)}(1)$,则

$$\hat{x}^{(1)}(k+1)=\left[x^{(0)}(1)-\frac{b}{a}\right]\mathrm{e}^{-ak}+\frac{b}{a}\quad(k=1,\ 2,\ \cdots,\ n)$$

(4) 还原值:

$$\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k)$$

上式即为预测方程。

有关建模的问题说明如下:

(1) 定原始序列 $X^{(0)}$ 中的数据不一定要全部用来建模,对原始数据的取舍不同,可得模型不同,即 a 和 b 不同。

(2) 模型的数据取舍应保证建模序列等时距、相连,不得有跳跃出现。

(3) 一般建模数据序列应当由最新的数据及其相邻数据构成,当再出现新数据时,可采用两种处理:一是将新信息加入原始序列中,重估参数;二是去掉原始序列中最老的一个数据,再加上最新的数据,所形成的序列和原序列维数相等,再重估参数。

7.3.2 GM(1, 1)模型检验

GM(1, 1)模型的检验分为三个方面:残差检验、关联度检验、后验差检验。

7.3.2.1 残差检验

残差大小检验,即对模型值和实际值的残差进行逐点检验。首先按模型计算

$\hat{x}^{(1)}(i+1)$，将 $\hat{x}^{(1)}(i+1)$ 累减生成 $\hat{x}^{(0)}(i)$，最后计算原始序列 $x^{(0)}(i)$与 $\hat{x}^{(0)}(i)$的绝对残差序列：

$$\Delta^{(0)} = \{\Delta^{(0)}(i),\ i=1,\ 2,\ \cdots,\ n\},\ \Delta^{(0)}(i) = |\ x^{(0)}(i) - \hat{x}^{(0)}(i)\ |$$

及相对残差序列：

$$\phi = \{\phi,\ i=1,\ 2,\ \cdots,\ n\},\ \phi_i = \left[\frac{\Delta^{(0)}(i)}{x^{(0)}(i)}\right]\%\ \cdot$$

并计算平均相对残差：

$$\bar{\phi} = \frac{1}{n}\sum_{i=1}^{n}\phi_i$$

给定 α，当 $\bar{\phi} < \alpha$，且 $\phi_n < \alpha$ 成立时，称模型为残差合格模型。

7.3.2.2　关联度检验

关联度检验，即通过考察模型曲线和建模序列曲线的相似程度进行检验。按前面所述的关联度计算方法，计算出 $\hat{x}^{(0)}(i)$与原始序列 $\hat{x}^{(0)}(i)$的关联系数，然后算出关联度，根据经验，关联度大于 0.6 便是满意的。

7.3.2.3　后验差检验

后验差检验，即对残差分布的统计特性进行检验。

(1) 计算出原始序列的平均值：

$$\bar{x}^{(0)} = \frac{1}{n}\sum_{i=1}^{n}x^{(0)}(i)$$

(2) 计算原始序列 $X^{(0)}$ 的均方差：

$$S_1 = \left(\frac{\sum_{i=1}^{n}[x^{(0)}(i) - \bar{x}^{(0)}]^2}{n-1}\right)^{1/2}$$

(3) 计算残差的均值：

$$\bar{\Delta} = \frac{1}{n}\sum_{i=1}^{n}\Delta^{(0)}(i)$$

(4) 计算残差的均方差：

$$S_2 = \left(\frac{\sum_{i=0}^{n}[\Delta^{(0)}(k) - \bar{\Delta}]^2}{n-1}\right)^{1/2}$$

(5) 计算方差比 C：

$$C=\frac{S_1}{S_2}$$

(6) 计算小残差概率：

$$P=P\{|\Delta^{(0)}(i)-\bar{\Delta}|<0.6745S_1\}$$

令 $S_0=0.6745S_1$，$e_i=|\Delta^{(0)}(i)-\bar{\Delta}|$，即 $P=P\{e_i<S_0\}$。

若对于给定 $C_0>0$，当 $C<C_0$ 时，称模型为均方差比合格模型；如对给定的 $P_0>0$，当 $P>P_0$ 时，称模型为小残差概率合格模型。后验差检验判别参照表如表 7-1 所示。

表 7-1 后验差检验判别参照表

P	C	模型精度
>0.95	<0.70	优
>0.80	<0.5	合格
>0.70	<0.65	勉强合格
<0.70	>0.65	不合格

若相对残差、关联度、后验差检验在允许的范围内，则可以用所建的模型进行预测，否则应进行残差修正。

7.3.3 GM(1, 1)模型应用实例

【例 7-1】 某散货码头 2007—2012 年的吞吐量如表 7-2 所示，试建立 GM(1, 1)预测模型，并预测 2013 年的产品销售额。

表 7-2 某散货码头 2007—2012 年的吞吐量 单位：百万吨

年 份	2007	2008	2009	2010	2011	2012
吞吐量	2.67	3.13	3.25	3.36	3.56	3.72

【解】 设 $X^{(0)}(k)=\{2.67, 3.13, 3.25, 3.36, 3.56, 3.72\}$

第 1 步：构造累加生成序列：

$$X^{(1)}(k)=\{2.67, 5.80, 9.05, 12.41, 15.97, 19.69\}$$

第 2 步：构造数据矩阵 $\boldsymbol{B}$ 和数据向量 $\boldsymbol{Y}_n$：

$$\boldsymbol{B}=\begin{bmatrix}-\frac{1}{2}[x^{(1)}(1)+x^{(1)}(2)] & 1\\ -\frac{1}{2}[x^{(1)}(2)+x^{(1)}(3)] & 1\\ -\frac{1}{2}[x^{(1)}(3)+x^{(1)}(4)] & 1\\ -\frac{1}{2}[x^{(1)}(4)+x^{(1)}(5)] & 1\\ -\frac{1}{2}[x^{(1)}(5)+x^{(1)}(6)] & 1\end{bmatrix}=\begin{bmatrix}-4.235 & 1\\ -7.425 & 1\\ -10.73 & 1\\ -14.19 & 1\\ -17.83 & 1\end{bmatrix}$$

$$\boldsymbol{Y}_n=\begin{bmatrix}x^{(0)}(2)\\ x^{(0)}(3)\\ x^{(0)}(4)\\ x^{(0)}(5)\\ x^{(0)}(6)\end{bmatrix}=\begin{bmatrix}3.13\\ 3.25\\ 3.36\\ 3.56\\ 3.72\end{bmatrix}$$

第 3 步：计算 $\hat{\boldsymbol{\alpha}}=\begin{bmatrix}a\\ b\end{bmatrix}=(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{Y}_n$：

$$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}=\begin{bmatrix}707.46375 & -54.41\\ -54.41 & 5\end{bmatrix}$$

$$(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}=\begin{bmatrix}0.008667 & 0.094319\\ 0.094319 & 1.226382\end{bmatrix}$$

$$\hat{\boldsymbol{\alpha}}=(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{Y}_n=\begin{bmatrix}-0.043879\\ 2.925663\end{bmatrix}$$

第 4 步：得出预测模型：

$$\frac{\mathrm{d}x^{(1)}}{\mathrm{d}t}-0.043879x^{(1)}=2.925663$$

$$\hat{x}^{(1)}(k+1)=69.3457\mathrm{e}^{0.043879k}-66.6757$$

$$\left(x^{(0)}(1)=2.67;\ \frac{b}{a}=-66.6757\right)$$

第 5 步：残差检验：

(1) 根据预测公式，计算 $\hat{x}^{(1)}(k)$，得

$$\hat{x}^{(1)}(k)=\{2.67,\ 5.78,\ 9.03,\ 12.43,\ 15.96,\ 19.68,\ 19.69\}\quad (k=0,\ 1,\ \cdots,\ 6)$$

(2) 累减生成 $\hat{x}^{(0)}(k)$序列，$k=0, 1, \cdots, 6$：

$$\hat{x}^{(1)}(k)=\{2.67, 3.11, 3.25, 3.40, 3.54, 3.71\}$$

原始序列：$X^{(0)}(k)=\{2.67, 3.13, 3.25, 3.36, 3.56, 3.72\}$。

(3) 计算绝对残差和相对残差序列：

绝对残差序列：$\Delta^{(0)}=\{0, 0.02, 0, 0.04, 0.02, 0.01\}$

相对残差序列：$\varphi=\{0, 0.64\%, 0, 1.19\%, 0.56\%, 0.27\%\}$

相对残差不超过 1.19%，模型精确度高。

第 6 步：进行关联度检验：

(1) 计算序列 $x^{(0)}$ 与 $\hat{x}^{(0)}$ 的绝对残差序列 $\Delta^{(0)}(k)$

$$\Delta^{(0)}=\{0, 0.02, 0, 0.04, 0.02, 0.01\}$$

$$\min\{\Delta^{(0)}(k)\}=\min\{0, 0.02, 0, 0.04, 0.02, 0.01\}=0$$

$$\max\{\Delta^{(0)}(k)\}=\max\{0, 0.02, 0, 0.04, 0.02, 0.01\}=0.04$$

(2) 计算关联系数。

由于只有两个序列(即一个参考序列，一个被比较序列)，故不再寻求第二级最小差和最大差。

$$\eta(k)=\frac{\min\{\Delta(k)\}+P\max\{\Delta(k)\}}{\Delta(k)+P\max\{\Delta(k)\}} \quad (k=1, \cdots, 6, P=0.5)$$

求得 $\eta(k)=\{1, 0.5, 1, 0.33, 0.5, 0.67\}$。

(3) 计算关联度。

$$r_i=\frac{1}{n}\sum_{k=1}^{n}\eta_i(k)=0.67$$

$r=0.67$ 是满足 $P=0.5$ 时的检验准则 $r>0.6$ 的。

第 7 步：后验差检验：

(1) 计算：

$$\bar{x}^{(0)}=\frac{1}{6}[2.67+3.13+3.25+3.36+3.56+3.72]=3.28$$

(2) 计算：$X^{(0)}$序列的均方差：

$$S_1=\left(\frac{\sum[x^{(0)}(k)-\bar{x}^{(0)}]^2}{n-1}\right)^{1/2}=0.3671$$

(3) 计算残差的均值：

$$\bar{\Delta}=\frac{1}{6}[\Delta(k)]=0.015$$

(4) 计算残差均方差：

$$S_1=\left\{\frac{\sum[\Delta(k)-\bar{\Delta}]^2}{n-1}\right\}^{1/2}=0.0152$$

(5) 计算 C：

$$C=\frac{S_1}{S_2}=0.0152/0.3671=0.0414$$

(6) 计算小残差概率：

$$S_0=0.6745\times0.3671=0.2746$$

$$e_k=|\Delta(k)-\bar{\Delta}|=\{0.15,\ 0.005,\ 0.015,\ 0.025,\ 0.005,\ 0.005\}$$

所有 e_i 都小于 S_0，故小残差概率 $P\{e_i<S_0\}=1$，而同时 $C=0.0414<0.35$，故模型 $\hat{x}^{(1)}(k+1)=69.3457\mathrm{e}^{0.043879k}-66.6757$ 合格。

第 8 步：预测：

$$k=7,\ x^{(0)}(8)=x^{(1)}(8)-x^{(1)}(7)=4.23$$

即 2013 年的吞吐量预测值为 4.23 亿元。

7.3.4 GM(1, 1)残差模型

当原始数据序列 $X^{(1)}$ 建立的 GM(1, 1)模型检验不合格时，可以用 GM(1, 1)残差模型来修正。如果原始序列建立的 GM(1, 1)模型不够精确，也可以用 GM(1, 1)残差模型来提高精度。

若用原始序列 $X^{(1)}$ 建立的 GM(1, 1)模型

$$\hat{x}^{(1)}(i+1)=\left[x^{(0)}(1)-\frac{b}{a}\right]\mathrm{e}^{-ai}+\frac{b}{a}$$

可获得生成序列 $X^{(1)}$ 的预测值，定义残差序列 $e^{(0)}(j)=x^{(1)}(j)-\hat{x}^{(1)}(j)$。若取 $j=i,\ i+1,\ \cdots,\ n$，则相应的残差序列为

$$e^{(0)}(k)=\{e^{(0)}(1),\ e^{(0)}(2),\ \cdots,\ e^{(0)}(n)\}$$

计算其生成序列 $e^{(1)}(k)$，并据此建立相应的 GM(1, 1)模型

$$\hat{e}^{(1)}(i+1)=\left[e^{(0)}(1)-\frac{b_e}{a_e}\right]e^{-a_e k}+\frac{b_e}{a_e}$$

得修正模型

$$x^{(1)}(k+1)=\left[x^{(0)}(1)-\frac{b}{a}\right]e^{-ak}+\frac{b}{a}+\delta(k-i)(-a_e)\left[e^{(0)}(1)-\frac{b_e}{a_e}\right]e^{-a_e k} \tag{7-5}$$

式中，$\delta(k-i)=\begin{cases}1 & k\geqslant i\\ 0 & k\leqslant i\end{cases}$为修正参数。

应用此模型时要考虑：

(1) 一般不是使用全部残差数据来建立模型，而只是利用部分残差。

(2) 修正模型所代表的是差分微分方程，其修正作用与 $\delta(k-i)$中的 i 的取值有关。

7.3.5 GM(1, N)模型

如果考虑的系统由若干个相互影响的因素组成，设 $X_1^{(0)}=\{x_1^{(0)}(1),x_1^{(0)}(2),\cdots,x_1^{(0)}(n)\}$ 为系统特征数据序列，而

$$X_2^{(0)}=\{x_2^{(0)}(1),x_2^{(0)}(2),\cdots,x_2^{(0)}(n)\}$$

$$\cdots\cdots$$

$$X_N^{(0)}=\{x_N^{(0)}(1),x_N^{(0)}(2),\cdots,x_N^{(0)}(n)\}$$

为相关因素序列。$X_i^{(1)}$ 为 $X_i^{(0)}$ 的 1—AGO 序列 ($i=1,2,\cdots,N$)，$z_1^{(1)}$ 为 $X_1^{(1)}$ 的紧邻生成序列，则称

$$x_1^{(0)}(k)+az_1^{(1)}(k)=\sum_{i=2}^{N}b_i x_i^{(1)}(k) \tag{7-6}$$

为 GM(1, N)灰色微分方程。

定义：

$$\hat{\boldsymbol{\alpha}}=[a\quad b_2\quad\cdots\quad b_N]^{\mathrm{T}}$$

为 GM(1, N)灰色微分方程的参数列，根据最小二乘法可以得出：

$$\hat{\boldsymbol{\alpha}}=(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{Y}$$

式中，

$$\boldsymbol{B}=\begin{bmatrix}-z_1^{(1)}(2) & x_2^{(1)}(2) & \cdots & x_N^{(1)}(2)\\ -z_1^{(1)}(3) & x_2^{(1)}(3) & \cdots & x_N^{(1)}(3)\\ & \cdots\cdots & & \\ -z_1^{(1)}(n) & x_2^{(1)}(n) & \cdots & x_N^{(1)}(n)\end{bmatrix}$$

$$\boldsymbol{Y}=[x_1^{(0)}(2)\quad x_1^{(0)}(3)\quad \cdots\quad x_1^{(0)}(n)]^{\mathrm{T}}$$

称

$$\frac{\mathrm{d}x_1^{(1)}}{\mathrm{d}t}+ax_1^{(1)}=b_2x_2^{(1)}+b_3x_3^{(1)}+\cdots+b_Nx_N^{(1)} \tag{7-7}$$

为 GM(1, N)灰色微分方程(7-6)的白化方程，也称影子方程。

于是，我们有：

(1) 白化方程(7-7)的解为

$$\begin{aligned}x_1^{(1)}(t)&=\mathrm{e}^{-at}\left[\sum_{i=2}^{N}\int b_i x_i^{(1)}(t)\mathrm{e}^{at}\mathrm{d}t+x_1^{(1)}(0)-\sum_{i=2}^{N}\int b_i x_i^{(1)}(0)\mathrm{d}t\right]\\&=\mathrm{e}^{-at}\left[x_1^{(1)}(0)-t\sum_{i=2}^{N}b_i x_i^{(1)}(0)+\sum_{i=2}^{N}\int b_i x_i^{(1)}(t)\mathrm{e}^{at}\mathrm{d}t\right]\end{aligned}$$

(2) 当 $X_1^{(1)}(i=1,2,\cdots,N)$ 变化幅度很小时，可视 $\sum_{i=2}^{N}b_i x_i^{(1)}(k)$ 为灰常量，这样 GM(1, N)灰色微分方程(7-6)近似时间响应式为

$$\hat{x}_1^{(1)}(k+1)=\left[x_1^{(1)}(0)-\frac{1}{a}\sum_{i=2}^{N}b_i x_i^{(1)}(k+1)\right]\mathrm{e}^{-ak}+\frac{1}{a}\sum_{i=2}^{N}b_i x_i^{(1)}(k+1) \tag{7-8}$$

式中，$x_1^{(1)}(0)$取为 $x_1^{(0)}(1)$。

(3) 累减还原式为

$$\hat{x}_1^{(0)}(k+1)=\hat{x}_1^{(1)}(k+1)-\hat{x}_1^{(1)}(k) \tag{7-9}$$

灰色系统建模的基本思路可以概括为以下几点：

(1) 建立模型常用的数据有以下几种：① 科学实验数据；② 经验数据；③ 生产数据；④ 决策数据。

(2) 序列生成数据是建立灰色模型的基础数据。

(3) 一般非负序列累加生成后，得到准光滑序列，对于满足光滑条件的序列，即可建立 GM 微分模型。

(4) 模型精度可以通过不同的灰数生成方式，数据的取舍，序列的调整、修正

以及不同级别的残差 GM(1, 1)模型补充得到提高。

(5) 灰色系统理论采用残差检验、关联度检验、后验差检验三种方法检验、判断模型精度。

7.4 灾变预测

灰色灾变预测的任务是给出下一个或几个异常值出现的时刻，以便人们提起防备，采取对策，减少损失。作为灰色预测模型的应用，以下简要介绍灰色灾变预测的原理和方法。

设原始序列为 $X=\{x(1), x(2), \cdots, x(n)\}$。给定上限异常值(灾变值)$\zeta$，成为 X 的子序列

$$\begin{aligned} X_\zeta &= \{x[q(1)], x[q(2)], \cdots, x[q(m)]\} \\ &= \{x[q(i)] \mid x[q(1)] \geqslant \zeta, i=1, 2, \cdots, m\} \end{aligned}$$

为上灾变序列。

设原始序列为 $X=\{x(1), x(2), \cdots, x(n)\}$。给定下限异常值(灾变值)$\xi$，成为 X 的子序列

$$\begin{aligned} X_\xi &= \{x[q(1)], x[q(2)], \cdots, x[q(l)]\} \\ &= \{x[q(i)] \mid x[q(1)] \leqslant \xi, i=1, 2, \cdots, l\} \end{aligned}$$

为下灾变序列。

设 X 为原始序列

$$X_\zeta = \{x[q(1)], x[q(2)], \cdots, x[q(m)]\} \subset X$$

为灾变序列，则称

$$Q^{(0)} = \{q(1), q(2), \cdots, q(m)\}$$

为灾变日期序列。

设 $Q^{(0)} = \{q(1), q(2), \cdots, q(m)\}$ 为灾变日期序列，其 1−AGO 序列为

$$Q^{(1)} = \{q^{(1)}(1), q^{(1)}(2), \cdots, q^{(1)}(m)\}$$

$Q^{(1)}$ 的紧邻生成序列为 $z^{(1)}$，则称 $q(k)+az^{(1)}(k)=b$ 为灾变 GM(1, 1)模型。

设 $\boldsymbol{\alpha}=[a \quad b]^{\mathrm{T}}$ 为灾变 GM(1, 1)模型参数序列的最小二乘估计，则灾变日期序列的 GM(1, 1)序号响应式为

$$\hat{q}^{(0)}(k+1)=\left(q(1)-\frac{b}{a}\right)\mathrm{e}^{-ak}+\frac{b}{a}$$

$$\hat{q}(k+1)=\hat{q}^{(1)}(k+1)-\hat{q}^{(1)}(k)$$

即

$$\hat{q}(k+1)=\left(q(1)-\frac{b}{a}\right)\mathrm{e}^{-ak}-\left(q(1)-\frac{b}{a}\right)\mathrm{e}^{-a(k-1)}=(1-\mathrm{e}^{a})\left(q(1)-\frac{b}{a}\right)\mathrm{e}^{-ak}$$

设 $X=\{x(1),\ x(2),\ \cdots,\ x(n)\}$ 为原始序列，n 为现在，给定异常值 ζ，相应的灾变日期序列：

$$Q^{(0)}=\{q(1),\ q(2),\ \cdots,\ q(m)\}$$

式中，$q(m)(\leqslant n)$ 为最近一次灾变发生的日期；$\hat{q}(m+1)$ 为下一次灾变的预测日期；对任意 $k>0$，称为 $\hat{q}(m+k)$ 未来第 k 次灾变的预测日期。

【例 7-2】 某地区平均降雨量（单位：mm）的原始数据为

$$\begin{aligned}X&=\{x(1),\ x(2),\ \cdots,\ x(24)\}\\&=\{386.6,\ 514.6,\ 434.1,\ 484.1,\ 647.0,\ 399.7,\\&\quad 498.7,\ 701.6,\ 254.5,\ 463.0,\ 745.0,\ 398.3,\\&\quad 554.5,\ 471.1,\ 384.5,\ 242.5,\ 671.7,\ 374.7,\\&\quad 458.9,\ 511.3,\ 530.8,\ 586.0,\ 387.1,\ 454.4\}\end{aligned}$$

规定年降水量 $\xi\leqslant 390$ mm 为旱灾年，试做旱灾预测。

【解】 首先做灾变映射。

按照 $x(t)\leqslant 390$ mm 为异常值，则有

$$\begin{aligned}X_{\xi}&=\{x[q(1)],\ x[q(2)],\ \cdots,\ x[q(6)]\}\\&=\{386.6,\ 254.5,\ 384.5,\ 242.5,\ 374.7,\ 387.1\}\\&=\{x(1),\ x(9),\ x(15),\ x(16),\ x(18),\ x(23)\}\end{aligned}$$

作异常值 $x[q(i)]$ 到出现灾变点 $q(i)$ 的映射 $Q^{(0)}$：$x[q(i)]\rightarrow q(i)$，得灾变日期序列 $Q^{(0)}$ 为

$$\begin{aligned}Q^{(0)}&=\{q(1),\ q(2),\ q(3),\ q(4),\ q(5),\ q(6)\}\\&=\{1,\ 9,\ 15,\ 16,\ 18,\ 23\}\end{aligned}$$

据此，对 $Q^{(0)}$ 建立灾变日期序列的 GM(1，1)模型。对 $Q^{(0)}$ 做一次累加生成，得

$$Q^{(1)}=\{q^{(1)}(1),\ q^{(1)}(2),\ q^{(1)}(3),\ q^{(1)}(4),\ q^{(1)}(5),\ q^{(1)}(6)\}$$

$= \{1, 10, 25, 41, 59, 82\}$

求得参数向量 $\hat{\boldsymbol{a}} = [a, b]^{\mathrm{T}} = [\boldsymbol{B}^{\mathrm{T}}, \boldsymbol{B}]^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{Y} = \begin{bmatrix} -0.188\,422 \\ 9.548\,72 \end{bmatrix}$。记 $Q^{(1)}$ 的紧邻生成序列为 $z^{(1)}$，于是，得灾变 GM(1, 1)为 $q(k) - 0.188\,422 z^{(1)}(k) = 9.548\,72$，灾变日期序列的 GM(1, 1)序列响应式为

$$\hat{q}^{(1)}(k+1) = \left(q(1) - \frac{b}{a}\right)e^{-ak} + \frac{b}{a}$$
$$= 51.677\,2e^{0.188\,422k} - 50.677\,3$$

从而

$$\hat{q}^{(1)}(k+1) = \hat{q}^{(1)}(k+1) - \hat{q}^{(1)}(k)$$
$$= 8.874\,78e^{0.188\,422k}$$

由此可得 $Q^{(0)}$ 的模拟序列：

$\hat{Q}^{(0)} = \{\hat{q}(k), k = 2, 3, 4, 5, 6\} = \{10.7, 12.9, 15.6, 18.8, 22.7\}$

由 $\Delta^{(0)}(k) = | x^{(0)}(k) - \hat{x}^{(0)}(k) |$，得绝对残差序列：

$\Delta^{(0)} = \{\Delta^{(0)}(k), k = 2, 3, 4, 5, 6\} = \{1.7, 2.1, 0.4, 0.8, 0.3\}$

及相对残差序列

$$\varphi = \left\{\varphi_i \mid \varphi_i = \left[\frac{\Delta^{(0)}(i)}{q(i)}\right], i = 2, \cdots, 6\right\} = \{0.19, 0.14, 0.025, 0.044, 0.013\}$$

平均相对残差：

$$\bar{\phi} = \frac{1}{5}\sum_{i=2}^{6}\phi_i = 0.08$$

小于 0.10，故可用

$$\hat{q}(k+1) = 8.874\,78e^{0.188\,422k}$$

进行预测。

$$\hat{q}(6+1) \approx 27,\ \hat{q}(7) - \hat{q}(6) \approx 5$$

即从最近一次旱灾发生的时间算起，5 年之后可能发生旱灾。

思考与练习

1. 名词解释：灰色，灰色系统，灰色预测，灰色灾变预测。

2. 灰色系统的特点有哪些?

3. 什么是 GM(1, 1)模型,其检验有哪些?

4. 设参考序列为

$$Y_0 = \{8, 8, 8, 16, 18, 24, 32\}$$

被比较序列为

$$Y_1 = \{10, 11, 16, 18.34, 20, 23.4, 30\}$$
$$Y_1 = \{5, 5.625, 5.375, 6.875, 8.125, 8.75\}$$

求其关联度。

5. 设有时间序列数据,如题表 7-1 所示:

题表 7-1

时 间	2009	2010	2011	2012	2013
k	1	2	3	4	5
$X^{(0)}(k)$	2.874	3.278	3.337	3.39	3.679

试建立 GM(1, 1)模型。

6. 设有时间序列数据,如题表 7-2 所示:

题表 7-2

时 间	2007	2008	2009	2010	2011	2012
k	1	2	3	4	5	6
$X^{(0)}(k)$	43.45	47.05	52.75	57.14	62.64	68.52

试预测 2014 年的数据。

8　预测方法的基本评价准则与比较

在预测工作中，一旦确定了预测目标，就必须选择适当的预测方法(技术)，而要选择预测方法，就需要对各种预测方法进行比较，选择其中的最佳方法。但是，各种预测方法都有自己的特点，每一个预测目标均有特定的预测要求，一种预测方法不可能达到所有的预测目标的预测要求。同样，一个预测目标的预测要求不可能所有的预测方法均能达到，预测要求与预测方法特点的不同，给严密确定预测方法的比较准则带来了许多问题，如何评价各种预测方法，以及如何根据确定的预测对象选用适合的预测方法，虽然现在还没有一套完善的准则，但人们在实践应用中还是逐渐建立了选择与评价预测方法的一些标准。本章将根据预测要求与预测方法特点的一般共同属性来确定预测方法的基本评价准则。

8.1　预测方法的基本评价准则

本节拟从5个方面来确定预测方法的基本评价准则，即预测误差、预测的时间和空间范围、预测的适用性、预测因子和资料、预测费用与预测效益。因为预测误差是选择预测方法时最重要的评价标准，而且是采用纯定量方法，故应做比较详细的研究。

8.1.1　预测误差

各种预测方法均或多或少地会产生预测误差，然而，人们总是希望预测误差降到最低限度。预测是一个极其复杂的过程，产生预测误差的因素甚多，为此，我们必须分析产生预测误差的因素，并设法消除或减少这些因素的影响。减少预测误差，可从两个方面考虑：其一，通过对各种预测方法产生预测误差的因素分析，来选择预测误差较小的预测方法，这是间接消除某些误差因素的一种方式。其二，在确定预测方法后，通过消除或减少某些误差因素来达到减少预测误差的目的，这可能使修正后的预测方法的精确度优于其他同预测对象的预测方法的精确度。因此，各种预测方法的预测误差比较是有条件的，即要在各种预测方法均进行了反复修正、反馈后，将各自产生的误差进行比较，才能确定最优精确度的预测方法。

8.1.1.1 产生预测误差的主要因素

1）预测因子（因素）的选择

不同的预测模型有不同的预测因子和因子的个数，当确定一个预测目标后，就要根据预测目标选择可能影响预测目标的各种因素。预测因子多，预测模型就相对复杂，误差就小，但工作人员、预测所需要的人力和财力就会增加；预测因子偏少，忽略了不应忽略的因素，则预测模型就相对简单，工作人员、预测所需要的人力和财力也相应减少，但误差往往增大。如果预测误差超过了规定的限度，则预测效果太差，就失去了预测的现实意义。因此，在经济、人力等客观条件允许下，尽量增加预测因子，使预测模型包含的信息量增大，这是缩小预测误差的途径之一。

2）不可预测的随机性

在预测中，有些预测因子本身是随机性因素，如洪水、干旱、战争、天灾人祸、经济危机等。我们必须尽量找出这些随机因素出现的规律性，并以某种人为假设，以便建立适合的预测模型，缩小预测误差。

3）观察与测量误差

预测因子所取的值不一定是客观存在的真值。仪器结构的不良，周围环境的改变，某些难以控制的偶然因素，粗枝大叶等原因均会导致观察或测量误差的产生，从而也就产生了预测误差。

4）数据资料的可靠性

准确而又全面地收集各种历史和现状资料，并加以整理和分析，是预测的基本步骤之一。如果缺乏数据资料、资料不全或者资料不准确，那么，整理和分析资料就会产生偏误，进而在选择预测技术、理论抽象、假设因素、建立模型条件等预测步骤中将导致连锁性偏误，最终产生预测误差。因此，在从事一项预测工作的期间，应该尽力利用各种方法，获得预测所需的准确资料。如果同类资料有不同的来源和出处，则应从多种来源分别收集、认真比较和核对，以便选出可靠与合理的资料，消除或减少由数据资料造成的预测误差。

5）建立模型的客观性

预测模型是不断修正、反馈后才建立的。在这个过程中存在着主观分析，判断与客观事实是否一致的问题。预测的每一个步骤都需要预测者进行分析和决断，不论是了解所要预测的客观过程的特性，还是选择预测的目标和影响预测的主要因素、预测资料的收集和整理、确定预测模型、选择最优预测模型、调整预测值等，都要求预测者进行分析和判断。预测过程的每一步若有分析或判断错误，将会导致其后的每一步分析和判断产生偏误，最终使预测值产生偏差。因此，预测工作者必须具有敏锐的观察能力，严密的逻辑思考能力和卓越的分析、判断能力。

6）方法选择的准确性

在预测工作的各个步骤中，均需要利用各种方法或技术，例如，确定预测因子的方法、收集原始预测资料的方法、对原始资料进行加工和整理的方法、选择预测技术的方法、建立预测模型的方法、选择最优预测模型的方法、检验预测模型的方法、自变量（外生变量）的预测方法、调整预测值的方法等。显而易见，对于给定的预测目标，以上种种方法的选择不同，得到的预测结果也是不同的。因此，预测者在条件允许的情况下，应选用产生误差最小的方法，使整个预测过程中产生的偏差最小。

8.1.1.2 预测误差的定义

设 n 个预测值为 $\hat{Y}_1, \hat{Y}_2, \hat{Y}_3, \cdots, \hat{Y}_n$；相应的 n 个实际值为 $Y_1, Y_2, Y_3, \cdots, Y_n$，把 Y_t 与 $\hat{Y}_t$ 的差记作 e_t，其中 $t=1, 2, \cdots, n$。

1）单个预测值的误差 e_t

它是指任一个实际值 Y_t 与相应的预测值 $\hat{Y}_t$ 的绝对误差。即

$$e_t = Y_t - \hat{Y}_t, \ t=1, 2, \cdots, n$$

当 $e_t=0$ 时，表示准确预测；$e_t>0$，说明 $\hat{Y}_t$ 为低估预测值；$e_t<0$ 时，表示 $\hat{Y}_t$ 为高估预测值。

2）单个预测值的相对误差 $\bar{e}_t$

它是绝对误差 e_t 与实际值的比值。即

$$\bar{e}_t = \frac{e_t}{Y_t} = \frac{Y_t - \hat{Y}_t}{Y_t}, \ t=1, 2, \cdots, n$$

当 $Y_t=1$ 时，$\bar{e}_t=e_t$；当 $Y_t>0$ 时，$\bar{e}_t$ 与 e_t 的意义相同；当 $Y_t=-1$ 时，$\bar{e}_t=-e_t$；当 $Y_t<0$ 时，$\bar{e}_t$ 与 e_t 的意义相反，即 $\bar{e}_t>0$ 表示 $\hat{Y}_t$ 是高估预测值，$\bar{e}_t<0$ 表示 $\hat{Y}_t$ 是低估预测值；$|\bar{e}_t|>1$，表明 $|e_t|>|Y_t|$，$|\bar{e}_t|$ 越大，则相对实际值 Y_t 的误差越大；$0<|\bar{e}_t|<1$，表明 $|e_t|<|Y_t|$，$|\bar{e}_t|$ 越小，则相对实际值 Y_t 的误差越小，$|\bar{e}_t|=0$，同样表示预测准确。

3）净预测误差 $\sum_{t=1}^{n} e_t$

它是 n 个预测值误差的代数和。即

$$\begin{aligned}
\sum_{t=1}^{n} e_t &= e_1 + e_2 + e_3 + \cdots + e_n \\
&= (Y_1 - \hat{Y}_1) + (Y_2 - \hat{Y}_2) + \cdots + (Y_n - \hat{Y}_n) \\
&= (Y_1 + Y_2 + \cdots + Y_n) - (\hat{Y}_1 + \hat{Y}_2 + \cdots + \hat{Y}_n)
\end{aligned}$$

$$= \sum_{t=1}^{n} Y_t - \sum_{t=1}^{n} \hat{Y}_t$$

当 $\sum_{t=1}^{n} e_t > 0$，表示全体实际值之和大于全体预测值之和，即总体预测值偏低。$\sum_{t=1}^{n} e_t < 0$，则表示全体实际值之和小于全体预测值之和，即总体预测值偏高。

4）预测的平均误差 MD

它是净预测误差的平均值。即

$$MD = \frac{1}{n}\sum_{t=1}^{n} e_t = \frac{1}{n}\sum_{t=1}^{n} (Y_t - \hat{Y}_t)$$

表示总体 e_i 的代表值(或一般水平)。

5）总预测绝对误差 $\sum_{t=1}^{n} |e_t|$

它是 n 个预测值误差的绝对值之和，即

$$\begin{aligned}\sum_{t=1}^{n} |e_t| &= |e_1| + |e_2| + \cdots + |e_n| \\ &= |Y_1 - \hat{Y}_1| + |Y_2 - \hat{Y}_2| + \cdots + |Y_n - \hat{Y}_n| \\ &= \sum_{t=1}^{n} |Y_t - \hat{Y}_t|\end{aligned}$$

6）平均绝对误差 MAD

它是一个预测值误差的绝对值的平均值。即

$$MAD = \frac{1}{n}\sum_{t=1}^{n} |e_t| = \frac{1}{n}\sum_{t=1}^{n} |Y_t - \hat{Y}_t|$$

7）相对误差的平均值 ARE

它是 n 个预测值相对误差的平均值。即

$$ARE = \frac{1}{n}\sum_{t=1}^{n} \bar{e}_t = \frac{1}{n}\sum_{t=1}^{n} \frac{Y_t - \hat{Y}_t}{Y_t}$$

8）相对误差绝对值的平均值 $AARE$

$$AARE = \frac{1}{n}\sum_{t=1}^{n} |\bar{e}_t| = \frac{1}{n}\sum_{t=1}^{n} \left|\frac{Y_t - \hat{Y}_t}{Y_t}\right|$$

9）预测误差的方差 S^2

定义有三种不同的形式：

$$S_1^2=\frac{1}{n}\sum_{t=1}^{n}e_t^2=\frac{1}{n}\sum_{t=1}^{n}(Y_t-\hat{Y}_t)^2$$

$$S_2^2=\frac{1}{n-1}\sum_{t=1}^{n}e_t^2=\frac{1}{n-1}\sum_{t=1}^{n}(Y_t-\hat{Y}_t)^2$$

$$S_3^2=\frac{1}{n-k}\sum_{t=1}^{n}e_t^2=\frac{1}{n-k}\sum_{t=1}^{n}(Y_t-\hat{Y}_t)^2 \text{（}n-k\text{ 为自由度的个数）}$$

10）预测误差的标准差 S

$$S=\sqrt{S^2}\text{，即 }S_1=\sqrt{S_1^2}\text{，}S_2=\sqrt{S_2^2}\text{，}S_3=\sqrt{S_3^2}$$

以上 10 种度量预测误差的方法，实际上是从不同角度分析预测误差的 10 种评价标准。因此，选择预测方法时，应多取几种评价标准，这样能从不同的角度反映各被选择的预测方法的预测误差大小。但是，一次只能使用一个评价标准进行比较，因为不同的度量方法得到的预测误差是不同的。例如，已知实际值为 9，其预测值为 8，另一个实际值为 1 000，其预测值为 1 001，则

$$e_9=Y_9-\hat{Y}_9=9-8=1,\ |e_9|=1,\ \bar{e}_9=\frac{e_9}{Y_9}=\frac{1}{9}=0.111$$

$$e_{1\,000}=Y_{1\,000}-\hat{Y}_{1\,000}=1\,000-1\,001=-1$$

$$|e_{1\,000}|=1,\ \bar{e}_{1\,000}=\frac{e_{1\,000}}{Y_{1\,000}}=\frac{-1}{1\,000}=-0.001$$

$$e_9=|e_9|,\ e_{1\,000}\neq|e_{1\,000}|,\ |e_9|>|\bar{e}_9|,\ |e_{1\,000}|>|\bar{e}_{1\,000}|$$

$|e_9|=1$，$|e_{1\,000}=1|$ 说明都有预测误差 1。$e_9=1$，$e_{1\,000}=-1$ 说明一个是预测偏低 1，一个是预测偏高 1。为了避免“正误差”和“负误差”在求平均值时相互抵消，掩盖了“正误差”和“负误差”同时产生的情况。我们引入了绝对误差 $|e_t|$，进而引入评价绝对误差 $MAD=\frac{1}{n}\sum_{t=1}^{n}|e_t|$，它不考虑预测值偏低或偏高，只考虑偏差的数值，故均取正号。然而 MAD 又有不足之处，如绝对值运算不便，反映大误差的灵敏度较低。我们引入了 S^2 的定义，从而弥补了这两点的不足。但是 e_t 又是其他预测误差定义的基础，不可缺少；S_1^2、S_2^2、S_3^2 的定义同时也为总体数值特征的计算提供了其特有的方法。

总之，不同度量方法其特点是不同的，预测者不应仅仅按照一种评价标准选择预测方法，必须注意的是，我们应该分清是预测方法本身造成的预测误差，还是其他因素。这有益于正确评价预测方法。

8.1.2 预测时间和空间范围

每一种预测方法都有其特定的预测时间和空间范围。选择预测方法的基本条件之一，就是使预测方法预测的时间和空间范围与决策者或应用者所规定的时间、空间范围相符。

预测的时间范围是指某项预测期限的长短，一般分为短期(1～3 个月)、近期(3 个月～2 年)、中期(2～5 年)和长期(5 年以上)4 种。当然这种划分不是绝对的，划分的实质性标志在于预测期内所有因素、经济结构是否发生变比，以及预测空间结构如何。

短期预测涉及的常常是一些确定性事项，随机因素的影响不大，内容繁杂琐碎；短期预测相对其他时间范围的预测比较简单，常用日或小时作为预测时间单位，一般用于微观的初、中级管理或经营的业务活动，有助于找出完成各种工作任务的最佳方法。

近期预测所涉及的问题，一般都受周期性和季节性影响很大，这与预测的时间单位有关，一般近期预测时间单位均用月。

中期预测时间单位一般用年，根据资料多少以及详细情况也可选用月为预测时间单位。如果以年为时间单位，则季节性因素在某年内与某年的下一年内对预测对象的影响基本上可以看作一致，因此，可把季节性因素作为一种确定性综合因素之一。但是，循环因素对中期预测影响很大，因为“循环”的周期是随机的；循环因素的变化会直接影响预测结果的精确性；循环因素将造成转折点的出现，而识别转折点对于决策者来说是非常重要和困难的，这就对中期预测方法的选择提出了更高的要求。

长期预测包含的不确定性因素比其他时间范围的预测要多得多，因此，长期预测不可能非常具体、详细。它主要用于确定长远的战略发展目标以及达到这种目标的途径等问题；预测因子一般均选用可控制的、对决策有重大影响的综合变量。鉴于长期预测模型或方法建立之时与预测对象实际到达的时刻相隔较长，因而，预测方法可随着时间的推进、信息的不断增加对以前的假定条件和某些控制变量进行必要的修正。

预测空间范围是指一个决策问题所包含的一个或几个甚至更多的预测对象。预测对象越多则预测空间范围越广。这些预测对象一般都是相互联系、相互影响的，有时可用方程组的形式表达出来。

预测空间范围还有一个密度概念，它包括两个方面的内容：一是预测因子的多少决定密度空间的维数，只有一个预测因子则密度空间的维数为一，如时间序列预测法、一元回归分析法均只有一维预测密度空间；二是预测因子所选用的单位大

小决定预测空间密度，如时间序列预测法中时间预测因子的单位有年、月、日、小时、分、秒。时间单位越小，对预测资料的要求就越多、越具体，其预测空间的密度就越大，这样也会使精确度越高。一般情况下，长期预测的密度空间维数要多于中、近、短期预测的密度空间维数。但是预测空间的密度则要小于中、近、短期预测的空间密度。总之，在选择预测方法时，应考虑各预测方法的预测空间的范围。这一评价预测方法的基本准则是一个比较综合的指标。

8.1.3 预测的适用性

预测方法的适用性是指实际应用这种方法的难易程度和实际应用的可能性。难易程度主要表现在：预测因子选择的多少，模型建立的难易，预测工作所需时间的长短。开展预测工作需要时间，而管理人员和决策者往往一提出预测课题，就希望马上得到结果，有时为了决策的需要，还常常限定在一定的时间范围内给出预测结果。因此，预测工作所需时间的长短也往往是选择预测方法的一个重要因素。

实际应用的可能性通常会出现两种情况：

一种情况是决策者和应用者对预测方法的偏爱程度一致，即应用者能代表决策者。在此情况下，选择预测方法主要依据预测方法的难易程度与应用者实际能力以及预测对象特性和预测方法特性是否一致相符，如预测对象的时间要求与预测方法所能预测的时间范围是否一致，预测对象有关的决策层次与详细程度是否与其预测方法的特性相对应。某部门、地区、企业管理人员或应用者的平均实施应用预测方法的能力与哪些预测方法的难易程度掌握相适应。如果某管理人员或应用者没有受过专门预测技术训练、缺乏相应的经济理论和数学、统计理论知识，则只能采用比较直观的、经验型的预测方法。

另一种情况是决策者和应用者对预测方法偏好不同，对于应用者来说，只有选择决策者偏爱的某些预测方法，才能使决策者接受。这时应用者必须了解决策者的政治环境、社会环境、经验和知识结构，特别是知识结构，它决定着决策者的主要思维方式。对于决策者来说，提出一系列预测目标要求，必须考虑具体的客观条件，同时明确应用者的能力以及预测工具所能达到的限度。预测方法从数据准备、模型建立、假设条件直至计算出预测结果，每一步骤中，客观限制的条件、应用者的各种能力以及预测工具的先进程度等都将直接影响到能否达到预测目标的要求。

8.1.4 预测因子和资料

从前面的章节中，我们知道要建立预测模型，首先要确定预测因子；为了方便选择预测方法，我们利用预测因子将各种方法建立的模型分为四类。

第一类模型是时间序列模型。

世上万事的存在发展都离不开时间，时间的有序性（有先后次序）能使人们有效地认识事物的存在和发展逻辑，并且可以预测未来事物的变化趋势。因此，最先考虑的预测因子应该是时间，大多数资料都是以历史数据形式出现，这就为时间序列模型的建立提供了方便。

第二类模型是因果关系模型。

它是以哲学中的范畴关系为基本理论，用数学形式来表达，即把某一变量看作是其他若干个变量的函数。这里，预测目标是果（结果），预测因子是因（结果发生的原因）。

第三类模型是统计模型。

预测因子确定等均是以概率论、数理统计原理为依据，具有严密的逻辑性，预测结果的可靠性强。但是，公理化的严密推导必然会附加许多假设和条件，这就会使模型结构复杂化。

第四类模型是除以上三种类型之外的其他模型。

总之，研究预测因子，有助于选择模型类别，而模型类别的确定，又为预测方法的选择范围的缩小提供了条件。

预测资料的性质和类型不仅是选择预测因子也是选择预测方法的一个极其重要的因素。因为有很多预测方法往往对不同类型的资料处理效果不同，尽管有些方法对数据类型的适应性较强，但并没有那些处理特定类型数据的预测方法有效。

客观生产管理过程中存在的资料数据，可先分为两大类：一类是横截面（或静态）数据，即确定某一时间的有关数据；另一类是纵截面（或动态）数据，即按时间顺序的排列有关数据。统计年鉴等书籍一般均以第二类方式给出数据。这样．要获得纵截面数据，就必须找出时期（点）相同数据重新整理、排列，预测对象与预测因子的关系。用函数图像分类，可根据以下三种特性进行：周期性、拟合性、综合性。

周期型数据是指在时间间隔内数据有规律地出现。例如，气候、季节，与季节有关的生产量、食品、服装等这些数据的变化周期是固定的，我们称这种周期型为季节型。还有一些数据变化的周期受很多随机因素的影响，并且同时还受其他许多因素的影响。例如经济危机、某些国际商品价格、某些商品的进口和出口量等，我们把这种不是在一个固定不变的时间间隔内反复出现，但又只有周期性的某些特性称为循环型。

拟合型数据是指数列能与某条直线或曲线方程图像拟合。拟合型又可分为直线型、非直线型、趋势型。趋势型是指某个变量在一定时间内有趋于增加或减少的趋势，如反映人们生活水平的某些指标、国民经济总产值、某些商品生产的成本等。

由于客观生产过程与经济环境密切联系，是一个非常复杂的过程，与某一经济

发展过程的数据对应的图像有时是几个或多个方程曲线的叠加，这种类型数据称为综合型。

对于水平型数据利用简单的移动平均法和指数平滑方法就可以很好地处理。对于较为复杂类型的数据，则需要利用高阶的指数平滑方法。回归分析方法有较强的适应能力，除严格的水平型数据之外，均可以处理。多元回归方法、经济计量模型方法、随机时间序列方法可以较好地处理季节型和循环型数据。由于定性预测方法不需要识别历史数据的类型，所以不存在适合某种类型数据的问题。

8.1.5 预测的费用与效益

预测的目的是为了减少决策损失，提前发现发展中的问题，找出发展的某些规律，使未来的效益达到最佳。虽然某些预测对象是属于政策和社会范畴，但是它们对经济效益的影响非常大，也可看作间接的经济效益。一般来说，使用预测方法的费用占所获经济效益的5%左右，有时也可高达10%以上。使用预测方法的费用也是衡量预测方法经济效果的重要参考量。使用各预测方法的费用比较，应该在同时达到了其他预测要求的条件下，相互进行费用比较。

预测方法的费用通常由以下几个方面组成：

(1) 调查费用。

(2) 收集及储存资料的费用。

(3) 研究预测模型的费用。

(4) 实际运用和维护的费用。

不同的预测方法，各种费用所占比重也不同。如定性预测方法一般花费的调查费用要占全体费用的较大比例。但在选择预测方法时，不能把费用的多少当作唯一因素。孤立地看重费用，反而不能提高预测工作的经济效益。

在进行预测时往往在给定一个预测课题的同时，也给定一笔费用，要求预测工作者在费用允许的范围内达到各项预测指标要求。这时，对于超出费用范围的预测方法，则其就没有被选择的权利。但是，对于费用在允许范围内的预测方法，单独的费用多少不能全面反映预测方法的优劣。最终要看预测的经济效益，来选择预测方法。

8.2 预测方法的比较

制定统一的预测方法比较标准，是比较困难的。因为比较事物都需要一定的范围和限制条件，而这些反过来又限制了事物的比较，如导弹与大炮、鱼翅与熊掌、手表与电视机、中国的长城与埃及的金字塔等。观察的场合不同、角度不同，对事

物优劣的评价也就可能不同，进而比较的结论也将会不同。这一节我们仅从两个方面来对预测方法进行比较。

8.2.1 预测方法类型的比较

8.2.1.1 定性预测方法类型与定量预测方法类型的比较

定性预测方法是对事物的性质作出描述。它的目的在于对预测事物有一个概括的了解，描述预测对象的发展趋势，分析判断某种现象出现的可能性。但是定性预测方法大都是以专家的意见和智慧为基础，通过集中、直观判断得出结果。对一些因素极为错综复杂、综合抽象的程度较大、牵涉社会心理因素较多的问题，定量预测模型的建立就很困难，预测对象与预测因子之间的关系也难以描述。这时常常利用定性预测方法来预测。专家们的智慧、综合理论和经验使他们有能力综合分析各种复杂因素，具有敏锐的洞察力，这些都有利于准确判断生产过程未来发展的性质，有利于转折点的预测和未曾出现过的新问题的未来发展性质的确定。定性预测方法简单，数学运算较少，有时甚至不需要计算，而是直接通过大脑进行预测。但是，定性预测方法也有一定的局限性，其主要表现在：

(1) 缺乏严密的逻辑推理、专家们的判断无统一的标准、仅凭各自的经验和知识，带有一定的主观性，意见有时难以统一。

(2) 预测结果的确定受主观性特别是权威性影响很大。

(3) 数量界限不清。

定量预测方法是利用数量测度来说明客观经济过程的未来发展。它能够给决策者提供确切的数量界限和科学依据。随着计算技术特别是电子计算机技术的高速发展和广泛应用，大量的信息资料的储存和提取以及复杂方程的求解等，均可在短期内完成，这就为定量预测方法的普遍使用提供了有力的工具，并克服了过去认为定量预测方法计算费时的缺点。定量预测方法具有以下优点：

(1) 定量预测方法是以数学、数理统计学等系统科学为理论基础，逻辑推理严密，能满足决策者的具体要求，能进行精确的科学预测。

(2) 定量预测方法大都是以数学模型综合多种因素描述客观经济现象的未来发展，可以假设许多条件，以修正某些随机因素和政策变化造成的预测偏差。

(3) 定量预测方法能够借助电子计算机，对客观经济过程中的未来发展进行数值模拟，为未来经济现象的实验研究，提供一个行之有效的途径。

(4) 定量预测方法对未来经济现象不但能够给出质的定性解释，而且能够给出量的确切描述。

当然，定量预测方法也有一些不足之处，如：方法繁多、对于决策者或管理人员来说，不可能一一掌握；很多方法适用范围有限等，特别是对转折点的预测不是

十分理想。这些都有待于进一步发展和完善。

在实际预测工作中，预测方法的选择并没有截然的定性和定量之分。定性预测方法需要定量处理、定量预测方法也需要进行定性分析，它们是相互补充、各有侧重的。

8.2.1.2　简单模型与复杂模型的比较

简单模型是指单方程模型，如时间序列模型、回归分析模型。复杂模型是指经济计量模型、投入产出模型等。

简单模型的预测空间范围相对于复杂模型来说要小得多，只有一个预测对象，有时甚至预测因子也只有一个。同时，简单模型也都是比较古典的模型，模型各种特点均研究得较透彻，不必再花时间、精力去研究，应用者可以像套数学公式一样去选择简单模型，如时间序列模型，只要找出历史数据与时间的函数关系或函数图像，即可选定拟合直线或曲线，建立预测模型。因此所花费的人力、物力和事件就少。

然而，简单模型预测复杂现象，由于忽略了很多因素，预测结果很难理想，这就要用到复杂模型。所谓复杂模型是指由多个有联系的单方程组所形成的模型。它不是仅分析某些特定序列的历史变化，而是根据相关理论、客观规律构造一系列的方程式；不仅包含经济因素，还包含非经济因素。利用这种模型进行预测，能够考虑非经济因素（政治、战争等）的影响。因为重大的变动往往是由外界因素造成，而模型考虑了这些影响，给出了数量测度，所以就可以研究不同的政策对未来的影响效果，从而可以为决策者提供不同政策方案。另外，我们在制订远景规划时，要使这种规划保持协调与平衡，既要考虑各种因素的直接影响，又要考虑各种因素的间接影响，如果没有复杂模型来描述，是很难达到目的的。

在实际工作中，复杂模型一般用于宏观预测，简单模型多用于微观预测。如果能够利用比较简单和费用较省的预测模型获得有一定精度要求的预测结果，就不必选择复杂的模型。复杂模型本身有许多问题（如自相关性、异方差等）需要进一步研究，因而在不少情况下还不如简单模型预测效果好。

8.2.2　相同预测对象的预测方法比较

在实际预测工作中，通常都要遇到同一个预测对象怎样选择最佳预测方法的问题。下面我们用评价预测方法的基本评价准则，对在相同预测对象情况下的预测方法进行比较。

决策者或各级管理人员制订的决策、规划和计划，都是相对某一确定的时间范围而言的，当给出预测目标后，预测时间和预测空间的范围也就基本上确定了。根据不同的预测时间，我们可选用短期、近期、中期或长期预测方法。短期预测最适

合的预测方法有：移动平均法、指数平滑法等。近期预测的预测方法有：回归预测法、趋势预测法等；适合中、长期的预测方法则有：回归分析法、复杂模型法（计量经济模型、投入产出模型）、德尔菲法、调查法等。在预测时间合适的条件下，我们应尽量选用预测空间范围广的预测方法。

预测方法的基本评价准则之间并不是相互独立的，它们存在着相互联系、相互依存、相互制约的关系。在选择预测方法时，如果只着重某一两个基本评价准则，那么被选择的预测方法就很难达到最佳。如果强调预测费用少则预测方法就不能太复杂，预测因子就需要尽量减少，这就限制了预测空间的范围、预测的精确性。也就是说，对预测费用的要求提高了预测空间范围的广度和密度，以及预测精确度的要求就要受到限制。同样，如果偏重于预测精确度的要求，预测费用就要增加，对预测资料的详细度、丰富度、准确度等的要求相应地也要增加，与预测目标有联系的因素，也将尽量纳入预测模型（即增加预测因子），这样预测模型就会复杂；反过来，由于模型建立的难度加大，又将使预测方法的广泛使用受到限制。

建立模型（方法）所需要的时间是预测的最基本要求之一，一旦给定了完成预测结果所需要的时间，我们就要从适合预测范围的各预测方法中，确定各预测方法的费时多少，从中再选出在费时范围之内的预测方法。预测方法费时的确定，可从资料收集、分析的时间多少、模型建立和修正的时间多少得到。如果预测工作者的水平较平均，则可用每人的费时相加来确定预测方法的总费时。当然，根据不同的条件和要求，计算预测方法的费时的多少会不一样，计算费时的方法也会有很大不同。如加权平均法，每人预测工作分段费时叠加法等这些都需要在实际中不断探索、完善。

预测方法有定性比较和定量比较。因为质总可以通过量的范围来表现。所以，定性比较总能直接或间接地化为定量的范围比较，通常定性比较实际上也是量的范围比较。我们可用 0～10 之间的数字表示不同方法拥有的等级水平。将预测方法不同量的指标进行等级分类，以便于比较和优选。至于各预测方法拥有的等级水平数字如何确定，这就要求预测者们对各种预测方法有深刻的了解，一般也有一定的规律可循。

预测方法的比较，是为了选择最优的预测方法，为决策者或管理人员提供科学依据和各种决策方案。因此，在可能的条件下，应当尽量同时采用几种方法进行预测，以便相互验证和补充。应该指出的是，每一经济过程中均有人的活动与意志参与；每一次预测方法的比较均加入了人的主观意念，这就要求预测工作者一切从实际出发，努力使自己的主观意识与客观实际相符合，使我们的预测工作做得更好。

思考与练习

1. 名词解释：预测误差。
2. 产生预测误差的主要因素有哪些？
3. 度量预测误差的方法有哪些？

下篇　决策理论与方法

9 决策技术概述

9.1 决策的基本问题

9.1.1 决策的定义

关于决策的定义，不同的学者理解不尽一致。西蒙对决策的解释较为宽泛，他的名言是“管理就是决策”。当代系统管理学家卡斯特(F. E. Kast)认为，决策就是进行判断和作出决定，即对两个以上的方案进行的考虑、权衡与选择，而行为是实现决策目标的过程，人们逼近目标靠的是不断进行决策和实现它们。当代另一位美国学者亨利·艾伯斯(Henry Embeth)则认为，决策有狭义和广义两种理解：从狭义方面说，决策就是在几种行为方案中做出抉择；从广义方面说，决策还包括在做出最后抉择前后所必须作出的一切活动。

在现代管理学中，决策就是人们为了实现特定的目标，在占有大量调研预测资料的基础上，运用科学的理论和方法，充分发挥人的智慧，系统地分析主客观条件，围绕既定目标拟订各种实施预选方案，从若干个有价值的目标方案中选择或实施一个最佳执行方案的活动。

决策有广义、狭义和最狭义三种解释。

决策的广义解释是把决策理解为一个过程。因为人们并不是突然确定行动方案的，而要经过提出问题、搜集资料、确定目标、拟订方案、分析评价、最终选定等一系列活动环节。而在方案选定之后，还要检查和监督它的执行情况，以便发现偏差，加以纠正。其中任何一个环节出了毛病，都会影响决策的效果。因此一个好的决策者，必然要懂得正确的决策程序，知道其中每个环节应当如何去做和要注意什么。

决策的狭义解释是把决策理解为仅仅是行动方案的最后选择，如我们常说的“拍板”。其实，判断、选择或“拍板”仅仅是决策全过程中的一个环节，如果没有“拍板”前的许多活动，“拍板”必然会成为主观武断的行为，决策也难免出乱子。

对决策这个概念还有一种最狭义的解释，即仅指在不确定条件下的方案选择。这类决策由于在面对客观环境中的不可控因素，要冒一定风险，因此有人认为，这种要担风险从而要靠决策者个人态度和决心来进行的决策才是决策。

随着社会的不断发展和人们对事物发展规律认识的深化,决策的制订和实施逐步由经验型向科学化发展。可见,科学化决策是经济、技术、社会发展的客观要求。

要进行决策,需要满足 4 个条件:① 决策总是为达到既定目的,离开了目的,就无所谓决策;② 决策总是要付诸实施,否则就是多余的活动;③ 决策总是在已确定的条件下寻求优化的目标和途径;④ 必须提出多个方案,否则就失去了优选的前提。

9.1.2 决策的基本要素

决策是一项系统工程,组成决策系统的基本因素如下:

1) 决策主体

决策是由人做出的,人是决策的主体,决策主体既可以是单个的个人.也可以是一个组织——由决策者所构成的系统。决策者进行决策的客观条件是他必须具有判断、选择和决断能力,承担决策后果的法定责任。

2) 决策目标

决策是围绕着目标展开的,决策的开端是确定目标,终端是实现目标。决策目标既体现了决策主体的主观意志,又反映了客观现实,没有决策目标就没有决策。

3) 决策对象

决策对象是决策的客体。决策对象涉及的领域十分广泛,可以包括人类活动的各个方面。决策对象具有一个共同点:人可以对决策对象施加影响。凡是人的行为不能施加影响的事物,不能作为决策对象。

4) 决策环境

决策环境是指相对于主体、构成主体存在条件的物质实体或社会文化要素。决策不是在一个孤立的封闭系统中进行的,而是依存于一定环境,同环境进行物质、能量和信息的交换。决策系统与环境构成一个密不可分的整体,它们之间相互影响、相互制约、息息相关。

9.1.3 决策的特点与类型

9.1.3.1 决策的特点

决策是一门方法论科学,具有以下基本特点:

1) 未来性

决策同预测一样是指向未来的。决策产生并存在于行动以前,决策关心的是用什么样的方式去实现将来的目标,因此决策将预测作为基础和依据,立足现在,把握未来。

2）不确定性

决策面对的是不确定性，一个不确定性的问题必然包含一定的风险。正是由于未来的不确定性所包含的风险才使决策越来越受到人们的重视和关注，由此人们把风险和决策常常联系在一起称作“风险决策”。事实上，如果一个问题是确定的，其结果是不容置疑的，那么这个问题就无需人们进行决策分析。

3）可选择性

决策面对的问题要有可选择性。决策是从通向未来的多种途径中做出明智选择的方法论科学。从某种意义上来说，决策的过程就是如何选择的过程。决策或选择的过程既体现科学手段、科学的理论和方法的客观标准，又体现了决策者的社会经验、学识水平、判断能力的主观标准。准确的信息是做出正确决策的基础。

4）系统性

一个决策问题应建立在某一决策系统内，具有可操作性和可实践性。决策应是在系统内谋求决策目标、内部条件以及外部环境之间的一种动态平衡。不同时点上的决策相互影响，具有一定的后效性，甚至形成动态联系。因此，决策应从整体上谋求最优或令人满意的行为措施。

9.1.3.2 决策的类型

1）按决策的范围可划分为宏观决策和微观决策

宏观决策是指对宏观经济活动的决策。宏观经济是以社会经济总量作为研究对象的，因而宏观决策一般带有全局性和综合性。微观决策是指对微观经济活动的决策。微观经济是以单个经济单位为研究对象的，因而微观决策一般带有个体性。

2）按决策的规模可划分为战略决策、战术决策和日常决策

战略决策是解决全局性、长远性、战略性的重大问题的决策，一般多由高层决策者做出。战术决策是在战略性决策做出以后，对如何保证完成任务和解决问题的决策，如组织所需的人力、物力、财力、技术等资源，一般由中层决策者做出。日常决策是指按照战术决策的要求对完成日常业务问题所做的短期的、技术性的、业务性的决策，一般多由基层决策者做出。

3）按决策的时间性可划分为长期决策和短期决策

长期决策是有关经济活动今后发展方向的长远性、全局性的重大决策，又称长期战略决策，短期决策是实现长期战略目标所采取的短期的、策略性的决策。

4）按决策的问题是否重复出现可划分为程序化决策和非程序化决策

程序化决策是按原来规定的程序、处理原则和方法去解决经济、管理中经常重复出现的问题的决策，又称常规决策。一部分战术决策和绝大部分日常决策多属程序化决策。非程序化决策，是用来解决以往没有先例的、新出现的、一次性的问

题的决策，又称非常规决策。非程序化决策往往是有关重大战略问题的决策。由于这种决策对各种因素变化的把握无法用常规办法处理，除采用定量分析外，更要借助于定性研究，且受决策者的观念、经验、知识、洞察力和直觉等因素的影响比较大。

5）按决策的时态可划分为单项决策和序贯决策

单项决策所处理的问题是某个时点的状态或某个时期总的结果，决策工作一次只有一个决策，又称静态决策。序贯决策是考虑到时间先后的动态变化而做出的一系列相互关联的多个决策，也称动态决策；序贯决策可以一次全部做出，也可以随上一步决策的实施逐步做出下一步的决策。

6）按决策的条件可划分为确定型决策、风险型决策和非确定型决策

确定型决策是指决策所依据的条件是确定的，结果也是确定的，一个方案只有一种确定的结果，只要比较各个方案的结果即可做出选择的决策。风险型决策所依据的条件和结果是不确定的，但是条件和结果发生的概率是可以估计的。由于概率是估计的，存在着不可控的因素，结果就具有一定的风险。非确定型决策不仅条件和结果是不确定的，而且条件和结果发生的概率也是无法估计的，完全凭机遇而定。

7）按决策要达到的要求可划分为最佳决策和满意决策

最佳决策是指在最理想的条件下做出的决策，这种决策要求在决策时所有的可能方案都要考虑到，从中筛选出最优的方案，每个方案都要按照最好的条件来设计。在个别不复杂的情况下，有可能做出这种决策；但在一般情况下，由于信息、资源、科技水平、认识能力等主客观条件的限制，理想的最佳决策是不容易做出的。因此，最佳决策往往只是一种理论上的追求。满意决策是指在现实可能的条件下以执行的结果使人满意为标准的决策，满意决策的优劣取决于对现实条件的充分分析。在实际的决策活动中，人们多半是以满意为标准的。不过，在若干个满意的目标或方案中，应力求选其更优者。

9.2 决策分析的程序

决策是一个复杂的动态系统，运行有序与否直接影响着决策质量。要提高管理决策的准确性和可靠性必须遵循科学的决策程序。一般来说科学的决策分析程序必须经过以下 7 个步骤。

1）发现问题

发现问题是决策工作的出发点，是决策者的重要职责。决策是为了解决将要发生的问题，而深入了解问题的特征并及时、准确地了解问题发生的信息是确定目

标的前提。发现问题、认识问题、分析问题是解决问题的关键，也是决策的关键。

2）确定目标

决策目标的确定一般要明确、合理，符合效率、经济性原则。一个具体决策目标应尽可能地数量化，否则应有度量目标达到程度的具体标准，为方案评价及决策效果评价提供依据。

确定决策目标首先要进行市场调查和科学的预测，调查和预测就是要依据过去和现在的各种信息，收集、查询原始信息和过去信息，以便对未来的发展变化做出科学预测。确定目标这一阶段通常被称为信息收集阶段，只有在大量掌握信息的前提下，才有可能提供决策目标。

为使决策目标明确、合理必须注意以下几个方面的问题：

第一，决策目标的确定要符合实际。

第二，决策的目标必须明确具体。

第三，要区分目标的重要程度和主次顺序。

同时，现在决策往往会面临多项目标并存的情况，在某些情况下，由于客观条件的限制，多个决策目标的作用方向可能不尽一致，甚至出现矛盾和对立，这就给方案的制订增加了难度。可以在满足决策要求的情况下，对多个目标按照其相互关系加以取舍，削减重复目标，合并类似目标，把决策目标减少到最低限度。

3）拟订方案

拟订备选方案，就是寻找达到目标的有效途径。途径有效与否要经过比较才能鉴别，因此必须制订多种可供选择的方案。多种方案必须有本质的区别，不是只有细节的差异。在拟订备选方案时，要注意以下两个方面的问题：

（1）方案的详尽性。详尽性即所拟订的全部备选方案中应当把所有的可能方案包括无遗，如果拟订的全部备选方案中漏掉了某些可能方案，那么最后选择的方案就有可能不是最优的。

（2）方案的相互排斥性。即不同的备选方案之间必须相互排斥，不能同时执行各方案，只能从中选择一个最满意方案。比如方案甲的行动措施包含于方案乙中，如为增加利润提出一个方案是降低产品成本，另一个方案是缩减管理费用，这两个方案就不存在选择的意义。相互排斥性是对信息加工处理的基本要求，要求避免方案的重复性和不可比性。

拟订备选方案可分为寻找和设计两个阶段：

（1）寻找备选方案。一般要依据大量的过去信息、固定信息，或根据以往的经验来拟订可供选择的方案。为此，必须进行大量的调查研究：一是要摸清情况，有目的地对大量统计数据进行分析处理，去粗取精，编制出简明扼要的表格与资料，提供给智囊系统和决策系统。关键性基础数据必须准确可靠，对动态数据的变化

情况和实际值要心中有数。二是广泛查阅、搜集与分析有关的国内外文献资料，尤其要了解国内外解决类似决策问题的方法、后果、经验与教训。除了积累文字情报以外，也应重视活情报的收集。三是为了决策科学化的需要，必须收集有关的信息，并加以处理、传送和使用，即要建立信息系统。四是由于决策所需要的条件和环境往往存在一些目前不能确定的因素，因此要在收集材料的基础上去粗取精、去伪存真、由此及彼、由表及里地分析和研究，并及时补充必要但尚短缺的资料，并根据决策目标的要求进行预测，提出可供抉择的多种方案，为比较对照提供前提。

(2) 设计阶段。需要对大胆寻找的方案"毛坯"进行精心设计、加工。如果说大胆寻找阶段需要勇于创新的精神和丰富的想象力，那么精心设计阶段则需要冷静的头脑和坚毅的精神。这是因为这里需要反复的计算、严格的论证和细致的推敲，还要经得起怀疑者的挑剔。设计阶段主要包括两项工作：一是对措施细节的确定，二是对方案后果的估计。

对措施细节的确定是一个完整的信息资料整理的过程，也是决策信息方案的制订过程。而对方案后果的估计实际上就是进行方案的评估工作，这里仍然需要预测信息。没有评估，就无从识别，最后的选择也无法进行。

4) 评定方案

这是整个决策过程中的关键环节、关键步骤。评定方案是一个比较、对照、分析和选择的过程。它的目的是从可供选择的多种可行性方案中选择一种最佳行动方案。

各备选方案评定的总原则是看哪个方案能更好地满足决策目标的要求。要评价各方案技术上的先进性、经济上的合理性和实现的可能性，计算直接经济效益和间接经济效益以及社会效益的大小，研究分析资源条件实现方案目标的可能性，同时还要估计各种备选方案潜在的问题以及可能带来的不良影响，从而做出全面、客观的综合评价。

对方案进行评估时，要根据目标来考核各个方案的费用和功效。首先要建立各方案的物理模型或数学模型，并求出各模型的解，对其结果进行评估。其次，要采用现代化的分析、评估、预测方法，对各种比较方案进行综合评价。一是运用定性、定量、定时的分析方法，评估各个比较方案的效能价值，预测决策的后果以及来自各阶层、各领域的反应。二是在评估的基础上，权衡、对比各个比较方案的利弊得失，并将各个比较方案按优先顺序排列，提出取舍意见，送交决策机构。

5) 选择方案

选择方案是决策程序中最为关键的环节，由决策系统完成。进行选择，就要比较可供选择方案的利弊，运用效能理论进行总体权衡、合理判断，然后选取其一，或综合成一，做出决策。

经过评估，对提供的各种方案做到了心中有数，再经过慎重考虑，便可进行选择。

对方案进行决策选择的常用方法有经验判断法和数学分析法。

(1) 经验判断法一般采用淘汰制，根据决策目标对备选方案进行筛选、淘汰，逐步缩小备选方案的选择范围。当余下的备选方案具有同等价值，而难以进一步选择时，可采取补充评选标准的办法，以表明某一方案优于其他方案。例如，各方案执行结果所获得的利润额差不多，则可以资金利用率和费用率作为补充标准，选择利润率高或费用率低的方案作为最优方案。

(2) 数学分析法是通过建立决策问题分析的数学模型，按照最优化决策准则进行方案优化选择，选出是优方案的方法。

6) 验证方案

方案选定后，有时需要进行局部实验，以验证其可靠性，通常称之为试点。试验实证是一个科学的步骤，必须科学地实施才具有意义。试验的选择必须是在全局情况中选择某些具有典型条件的点，并且严格按照所决策的方案实施，同时还必须有相同条件的一般对照组，这样才能从比较中得出科学的结论。如果试点成功，就可进入全面普遍实施阶段。如果不行，那就必须反馈回去，进行决策修正。由于决策是一个动态的依赖于时空变量的复杂随机函数，为了能客观地反映决策合理与否，还需要进行可靠性分析。

7) 方案的普遍实施

这是决策程序的最终阶段。由于通过上一阶段的试验验证，可靠程度一般是较高的。但是，在执行过程中仍可能发生一些以前没有考虑的偏离目标的情况。因此为了确保决策的实施，必须做到以下几个方面：

(1) 制订实施计划。

(2) 建立相应的组织。

(3) 做好宣传、动员工作。

(4) 要注意决策实施中信息的反馈。

9.3 决策方案的敏感性分析

敏感性分析是通过分析、预测项目主要因素发生变化时对相关评价指标的影响，从中找出敏感因素，并确定其影响程度，从而了解项目的风险因素及风险程度，考察项目承受风险的能力。敏感性分析分为单因素敏感性分析和多因素敏感性分析。

1) 单因素的敏感性分析

单因素的敏感性分析是假定其他因素不发生变化而分析某一种因素单独变动

所引起的项目相关评价指标的变动幅度或边界的方法。首先，要确定评价的对象，如净现值、总收益、总成本、投资回收期等。其次，确定敏感性因素，如产品的市场价格、项目寿命期等。然后根据指标和分析参数之间的关系，对每一参数的变动幅度及其影响进行分析。

2）多因素的敏感性分析

多因素的敏感性分析是指每次同时变动两个或两个以上参数，其他因素保持不变，观察和分析当这些因素发生变动时，对项目相关指标的影响。在进行多因素敏感性分析时，一次改变两个参数可以得到敏感性曲线，若三个参数同时变化，则可以得到一个敏感面。

3）敏感性分析的作用

（1）可用来对决策方案进行风险分析，考察决策方案的风险程度和承担风险的能力。决策方案的风险程度和承担风险的能力主要取决于风险因素的多少和决策方案对风险因素的敏感程度高低。

（2）可用来找出决策方案的风险因素，以便采取措施控制风险因素的变动范围，减少其变化量，从而降低决策的风险。例如，通过敏感性分析得知决策方案对经营成本很敏感，就可知经营成本是主要风险因素，我们就可加强成本管理，努力把成本控制在最小限度范围内，这样就可把由成本的不确定性引起的决策风险降低到最低程度。

（3）可用来进行多方案比较，从中选择风险最小的方案。一般说来，不同的决策方案对同一不确定性因素的敏感程度是不相同的，敏感程度小的方案即为风险小的方案。因此，如果几个方案的相关评价指标都达到了要求，且他们的效用基本相同，应取敏感性程度小的方案。

（4）可揭示决策方案的各不确定因素和决策方案效用之间的因果关系，找出影响决策方案经济效益的最主要因素，进一步提高与之相关的数据预测和估算的可靠程度，从而提高决策方案的评价质量。

4）敏感性分析的步骤

（1）选择具体评价指标为敏感性分析对象。在进行敏感性分析时，并不要求对所有的评价指标都进行分析，而是只选择最能反映项目效用的指标作为分析对象，而且不同特点的项目反映效用的指标也不完全相同。

（2）选择不确定因素作为敏感性分析变量。项目的不确定性因素一般有生产量、价格、主要原材料价格或动力价格、可变成本、固定资产投资、建设工期及外汇汇率，其中产量、价格、成本、投资等因素是最常被选择的变量。

（3）确定变量的变化范围。变量的变化范围一般是根据历史的统计资料和对市场的调查预测来进行估计，估计值可比历史统计资料和预测值略偏大。

(4) 计算由于各变量的变化引起敏感性分析对象变动的幅度。

(5) 通过比较找出项目敏感性因素,即风险因素,把各变量在其变化范围内变化引起的分析对象的变动幅度进行比较,引起分析对象变动幅度较大的变量即敏感性因素、业绩风险因素。

5) 敏感性分析的局限性

(1) 敏感性分析只是用以考核某些主要变量的变化对决策方案效果影响程度的简单分析手段,但不能估计风险程度和出现的可能性。也就是说,敏感性分析只能对风险进行定性评估,而不能对风险进行定量预测。

(2) 只有在多方案比较时,敏感性分析的结果才可成为决策方案取舍的依据。在单一方案情况下,敏感性分析的结果只用于对决策方案实施的风险进行评价,一般情况下,不能作为决策方案取舍的依据。

(3) 在运用敏感性分析对决策方案进行风险分析时,各不确定因素的变化方向和变化范围被认为是确定的,而实际上变化方向和变化范围也是不确定的,但敏感性分析却没有给出它们发生的概率。由此得出的有关决策方案风险的评价显然欠科学。

(4) 敏感性分析的工作量太大。一个决策方案的不确定性因素往往有许多,每个不确定性因素都要取几个变化值来分别计算它们引起的指标的变化幅度。另外,在进行这样的计算和分析时,往往要几个方案、几个项目综合考虑。这样,涉及的数据很多,计算的工作量十分大。例如,一个决策方案如果有 5 个敏感性因素,每个因素取 3 个变化值,那么就要做出 3^5,共 243 个敏感分析,这样的计算量和工作量过多。因此,敏感性分析不能乱用。

思考与练习

1. 名词解释:决策,敏感性分析。
2. 说明决策的基本要素。
3. 决策分析的程序是什么?
4. 敏感性分析有什么样的局限性?

10　确定型决策

10.1　确定型决策概述

10.1.1　确定型决策的概念

确定型决策是指当决策者面对决策问题进行抉择时，每一个抉择行动只能产生一个确定结果，决策者可以根据完全确定的情况，选择最满意的方案或最优行动。例如，某人得到一笔小奖金 200 元，他可以用这些钱买一份礼物送给父母，可以给儿子买他向往已久的玩具汽车，可以一家三口出去吃一顿，还可以为自己买一些资料。他做出一个决策，采用了以上的其中一条。比如买礼物送给父母，那么结果就是表示了孝心，这就是一个确定型决策。

确定型决策应具备的条件也就是应用确定型决策方法的条件具体有：① 存在决策者期望达到的一个决策目标；② 未来的状况，只存在一个确定的自然状态；③ 存在两个或两个以上的备选方案，供决策者选择；④ 每个备选方案在确定状态下的损益值可以计算出来，如港口企业生产规模的确定、港机零配件库存最佳规模的设计、进货批量的安排等问题。符合上述条件的预测，就可采用确定型决策方法。确定型决策方法主要有盈亏平衡分析法、线性规划法等。

10.1.2　确定型决策的基本思路

一般确定型决策可以用单纯优选决策法和模型优选的数学决策分析决策法来进行。

1）单纯优选法

单纯优选法是一种简单的决策方法。如果决策者遇到的是这样一类决策问题，其行动方案仅是有限个，而且掌握的数据资料也无需加工计算，就可以逐个比较直接选出最佳方案和最优行动，这种在确定情况下的决策就是单纯优选决策。如利率不同的多种渠道均可以筹措到一笔资金，在单一决策目标（即筹资成本最低）下，就可以选定利率最低的渠道去筹集资金。

2）模型优选法

确定型决策问题看起来似乎很简单，但是有的实际问题却相当复杂，有时可供

选择的方案很多，难以选择。例如：有 N 个产地、M 个销售地区的某种物质的运输问题，当 N、M 较大时，这就需要我们找出运费最少的方案。对于确定性决策的模型优选，一般常用的方法有：线性规划、非线性规划、整数规划、动态规划、确定性存储技术、网络分析技术等。

模型优选决策方法的基本思路是：

(1) 决策目标的设计。包括单一目标的决策和多目标的决策。在多目标决策问题中，还要区分各个目标之间的优先级顺序及重要程度。

(2) 确定型决策的约束条件的建立。有些确定型决策问题要实现指标的最大化或最小化是有一定限制条件的，如资源的限制等。这时，要得到最优方案，必须在满足约束条件的基础上进行。

(3) 求解确定型决策的优化解，即最优方案。

10.2 线性盈亏平衡决策模型

10.2.1 线性盈亏平衡分析假设条件

盈亏平衡分析是通过项目盈亏平衡点分析项目成本与收益平衡关系的一种方法(盈亏平衡点又称为保本点、盈亏临界点，是销售收入等于总成本费用、不盈不亏的临界状态)，分为线性盈亏平衡分析和非线性盈亏平衡分析。本书重点介绍线性盈亏平衡分析方面的内容。

所谓线性盈亏平衡分析，就是对企业总成本和总收益的变化作线性分析，目的在于掌握企业经营的亏损界限，确定企业的最优生产规模，使企业获得最大的经济效益，以做出合理的决策。

线性盈亏平衡分析假设条件有：

(1) 项目总销售收入和生产总成本费用是产品销售量的线性函数。

(2) 产品(服务)销售量和生产量相等(即各年产品全部售出)。

(3) 产品(服务)固定成本和单位售价在项目生产期内保持不变。

(4) 项目生产(提供)的是单一产品(服务)，如同时生产(提供)几种类似产品(服务)，则应把几种产品(服务)组合折算为一种产品(服务)。

(5) 各种数据应取正常生产年度(即达到设计生产能力的年份)的数据。

10.2.2 线性盈亏平衡决策模型

10.2.2.1 图解法

将固定成本线置于变动成本线之下，反映出固定成本不随产量变动而变动的

特性。如图 10-1 所示为销售收入线与成本费用线的交点为盈亏平衡点。当产量或销售收入超过这一盈亏平衡点，进入盈利区；反之，则为亏损区。在盈亏平衡点以下的项目应该放弃。

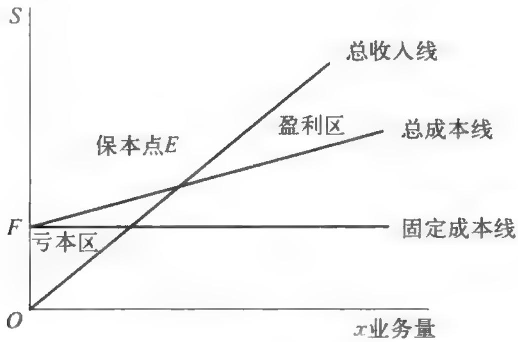


图 10-1 传统式保本图

10.2.2.2 解析法

解析法也称为公式法，通过在各个分析指标之间建立某种线性关系来分析在不确定条件下项目的经济价值。假设 TR 为企业的年总收入，TC 为年总成本，P 为单位售价，c 为单位可变成本，C_V为年总可变成本，Q 为年生产产量，C_F 为年固定成本，c_r 为年销售税金，a 为税率，则其基本的线性关系是

$$TR = PQ \tag{10-1}$$

$$c_r = PQa = TRa \tag{10-2}$$

$$TC = C_F + cQ \tag{10-3}$$

$$TR - c_r = TC \tag{10-4}$$

$$PQ(1-a) = C_F + cQ \tag{10-5}$$

其中，式(10-1)为年总收入 TR 是年产量(销售量)Q 与产品(服务)单价 P 的乘积；式(10-2)为年销售税金 c_r 是年总收入 TR 与税率 a 的乘积；式(10-3)为年总成本 TC 是固定成本 C_F 与可变成本 cQ 之和；式(10-4)为在盈亏平衡点处，年总收入 TR 是年总成本 TC 与销售税金 c_r 之和；式(10-5)为式(10-4)的各项内容的展开。

根据上述计算过程，线性盈亏平衡点的计算公式有 3 个：

(1) 以产(销)量表示的盈亏平衡点的计算公式。

$$Q^* = \frac{C_F}{P(1-a)-c} \tag{10-6}$$

式中，Q^* 为盈亏平衡点的产(销)量。

(2) 以生产能力利用率表示的盈亏平衡点计算公式。

生产能力利用率是实际年产量和设计年产量(设计生产能力)之比。设实际年产量为 Q，设计年产量为 Q_{max}，则当设计年产量充分利用的情况下，以下等式成立：

$$TR_{max} = PQ_{max} \tag{10-7}$$

$$C_{Vmax} = cQ_{max} \tag{10-8}$$

用 q 表示产能利用率，则

$$q = \frac{Q}{Q_{max}} \times 100\% \tag{10-9}$$

用 q^* 表示盈亏平衡点处的产能利用率，在用产量表示的盈亏平衡公式中两边分别除以设计年产量，则得到

$$q^* = \frac{C_F}{Q_{max} \times (1-a) - C_{Vmax}} \tag{10-10}$$

显然，盈亏平衡点的产量等于设计生产能力乘以生产能力利用率。

(3) 以销售额表示的盈亏平衡点计算公式。

设 TR^* 表示盈亏平衡点的销售额，在式(10-6)两边分别乘上销售单价，则能得到

$$TR^* = \frac{C_F}{1-a-c/P} \tag{10-11}$$

【例 10-1】 某港口投资项目建成后，在正常生产年份每提供 1 次港口服务可收入 800 元，预计年提供服务总量可达 10 万次，年固定成本额为 1 000 万元，单位变动成本为 350 元，销售税率为 15%，分别计算以生产能力利用率和年销售收入表示的盈亏平衡点。

【解】 (1) 根据式(10-10)可得

$$\begin{aligned} q^* &= \frac{C_F}{Q_{max} \times (1-a) - C_{Vmax}} \\ &= 1\,000 \div [800 \times 10 \times (1-0.15) - 350 \times 10] = 30\% \end{aligned}$$

(2) 根据式(10-11)可得

$$\begin{aligned} TR^* &= \frac{C_F}{1-a-c/P} \\ &= 1\,000 \div (1-0.15-350/800) = 1\,000 \div 0.412\,5 = 2\,424(\text{万元}) \end{aligned}$$

所以，该项目盈亏平衡点的生产能力利用率为 30%，盈亏平衡点的销售收入额约为 2 424 万元。

【例 10-2】 生产规模与购置选择的盈亏分析模型。某港口企业正在筹建一个新项目，提出 3 个方案：① 采用高度自动化设备，固定成本将较高，达到 800 万元，单位可变成本为 10 元；② 采用半自动化设备，固定成本为 600 万元，单位可变成本为 12 元；③ 采用非自动化设备，固定生产成本虽然较低为 400 万元，但单位可变成本却较高为 16 元。在此基础上试确定该项目的最佳建设方案。

【解】 经过分析，我们可以看到，决策的目标是在一定产量下取得最高的利润或达到生产成本最低。假设售价不变，利润的高低取决于生产规模，即在不同的生产规模下，采用的最佳方案将会有所不同。设年产量为 Q，则各方案的总成本分别为

$$TC_1 = 800 + 10Q$$
$$TC_2 = 600 + 12Q$$
$$TC_3 = 400 + 16Q$$

将这 3 条总成本线描绘在同一图中，形成总成本分析图，如图 10-2 所示。

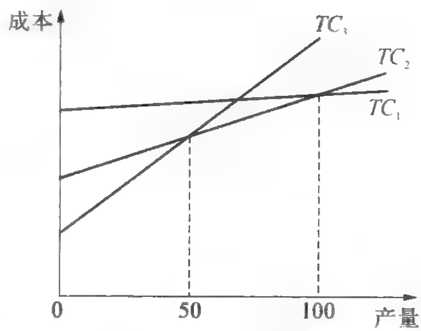


图 10-2　成本结构分析图

当生产规模为 50 万元时，$TC_2 = TC_3$；当生产规模为 100 万元时，$TC_2 = TC_1$。

当生产规模小于 50 万元时，第三方案的总成本最低，应采用第三方案。当生产规模大于 50 万元而小于 100 万元时，第二方案的总成本最低，应采用第二方案。当生产规模大于 100 万元时，第一方案的总成本最低，应采用第一方案。

10.2.3 盈亏平衡分析的作用

(1) 可用于对项目进行定性风险分析，考察项目承受风险的能力。

(2) 适用于进行多方案的比较和分析，在其他条件相同的情况下，盈亏平衡点值低的方案为最优方案。

(3) 可用于分析价格、产销量、成本等因素变化对项目盈利能力的影响，寻求提高盈利能力的途径。一般来说，价格和销量由市场来定，难以控制，所以降低成本是提高盈利能力的主要途径。

10.2.4 盈亏平衡分析的局限性

(1) 只能对项目风险进行定性分析,无法定量预测风险的大小。

(2) 由于它所假设的条件往往和实际情况有出入,加上没有考虑资金时间价值和项目整个寿命期内现金流量的变化,所以分析欠准确。

(3) 它只是分析了价格、产销量、成本等因素变化对盈利能力的影响,而不能确定盈利能力大小。

所以,在决策方案选择时,盈亏平衡决策分析还需要和其他方法结合使用,才能提高分析效果。

10.3 线性规划决策模型

在社会经济活动中,人们总是希望通过某种途径,追求可能达到的最佳结果,这就是优化问题。其基本思路是在满足一定的约束条件下,使预定的目标值达到最优。

线性规划最初是为解决二次大战中的后勤问题而产生的。自 1947 年丹茨格(G. B. Dantzig)提出求解线性规划的单纯型法后,线性规划的理论体系和计算方法日趋系统和完善。随着电子计算机的发展,线性规划已广泛应用于商业、工业和军事,例如人力资源规划、选址问题、库存管理、生产计划、投资分析、营销决策等,在港口管理领域的应用也十分广泛。

10.3.1 最大化问题

下面以一个简单的最大化问题为例,说明线性规划的思路、建模与求解方法。

【例 10-3】 港口机械厂生产优化问题。

某港口机械厂生产 A、B 两种产品。产品 A 与产品 B 在生产过程中均需使用原材料 1,其中每件所需消耗的原材料 1 的数量分别为 6 与 2。同时,产品 B 还需使用原材料 2,每件产品的消耗量为 1。此外,每生产一件产品 A 与一件产品 B 所需的劳动时间分别为 2 与 4。该厂可提供的两种原材料和劳动时间的数量是有限的。在第一个月初,该厂可提供的原材料 1 的数量为 1 800,原材料 2 的数量为 350,可提供的劳动时间为 1 600。该两种原材料的保存时间是一个月,也就是说第一个月用不完的原材料只能丢弃。经财务部门分析计算,产品 A 与 B 每件利润分别为 3 元与 8 元。而且根据市场调查得到的该两种产品的市场需求状况可以确定,按当前的定价可确保所有产品均能销售出去。问第一个月内产品 A 与产品 B 各应生产多少,可使总利润最大?

在上述问题中，目标是总利润的最大化，所要决策的变量是产品的产量，而产品的产量则受到可提供的原材料与劳动时间的约束，因此该问题可以用目标、决策变量和约束条件三个因素加以描述。实际上，所有的线性规划问题都包含这三个因素。现对这三个因素简单说明如下：

(1) 目标函数是指系统所追求的目标的数学描述。例如最大利润，最小成本等。

(2) 决策变量是指系统中有待确定的未知因素。例如决定企业经营目标的各产品的产量等。

(3) 约束条件是指实现系统目标的限制因素，它们限制了目标值所能达到的程度。例如原材料供应量、市场需求等。

下面对上述问题进行分析与求解。

【解】 本问题可用表 10－1 表示。

表 10－1 港口机械厂月生产安排

项　目	1 件产品 A	1 件产品 B	总　量
原材料 1	6	2	1 800
原材料 2	0	1	350
劳动时间	2	4	1 600
利　润	3	8	

(1) 决策变量：本问题的决策变量是第一个月产品 A 与产品 B 的产量。可设：

X 为第一个月产品 A 的产量(件)；Y 为第一个月产品 B 的产量(件)。X、Y 即为本问题的决策变量。

(2) 目标函数：本问题的目标函数是总利润最大。由于产品 A 与 B 每件利润分别为 3 元与 8 元，而其产量分别为 X 与 Y，所以总利润可计算如下：

$$\text{总利润} = 3X + 8Y(\text{元})$$

(3) 约束条件：本问题共有 4 个约束。第一个约束是原材料 1 的约束。每件产品 A 与产品 B 对原材料 1 的消耗量分别为 6 与 2，而两种产品的产量分别为 X 与 Y，所以该两种产品在第一个月对原材料 1 的总消耗量为 $6X+2Y$。由题意，原材料 1 的可提供量为 1 800。由此可得第一个约束如下：

$$6X + 2Y \leqslant 1\,800$$

第二个约束是原材料 2 的约束。由于只有产品 B 需消耗原材料 2，而且单位

产品 B 对原材料 2 的消耗量为 1，产品 B 的产量为 Y，所以原材料 2 的总消耗量为 Y。由题意，原材料 2 的可提供量为 350。由此可得第二个约束如下：

$$Y \leqslant 350$$

第三个约束是劳动时间的约束。由于每单位产品 A 与 B 对劳动时间的需要量分别为 2 与 4，而两种产品的产量分别为 X 与 Y，所以两种产品在第一个月所需的总劳动时间为 $2X+4Y$。由题意，劳动时间的可提供量为 1 600。由此可得第三个约束如下：

$$2X + 4Y \leqslant 1\,600$$

第四个约束是非负约束。由于产量不可能为负值，所以有：

$$X \geqslant 0,\ Y \geqslant 0$$

由上述分析可建立本问题的线性规划模型如下：

$$
\begin{aligned}
\text{o. b.}\quad & \max \quad F = 3X + 8Y \\
\text{s. t.}\quad & 6X + 2Y \leqslant 1\,800 \quad （原材料\ 1\ 约束） \\
& Y \leqslant 350 \quad （原材料\ 2\ 约束） \\
& 2X + 4Y \leqslant 1\,600 \quad （劳动时间约束） \\
& X \geqslant 0,\ Y \geqslant 0 \quad （非负约束）
\end{aligned}
$$

下一步就是要找出决策变量 X 与 Y 的值，使得在同时满足所有约束条件的前提下目标函数值达到最优，这就是线性规划的求解。

本章讨论的问题均为线性规划问题。所谓"线性"规划，是指如果目标函数是关于决策变量的线性函数，而且约束条件也都是关于决策变量的线性等式或线性不等式，则相应的规划问题就称为线性规划问题。本题中，目标函数 $3X+8Y$ 是关于决策变量 X 与 Y 的线性函数，而 4 个约束条件也都是关于决策变量 X 与 Y 的线性不等式，所以本问题是一个线性规划问题。

10.3.2 线性规划问题的图解法

10.3.2.1 可行域与最优解

在【例 10－3】中所要寻求的解是产品 A 与产品 B 的产量结合。实际上，给出产品 A 与产品 B 的任意一组产量组合，就可能得到该问题的一个解，因此可以得到无穷多个解，但是其中只有满足所有约束条件的解才是符合题意的。满足所有约束条件的解称为该线性规划问题的可行解，全体可行解组成的集合称为该线性规划问题的可行域。

其中，使得目标函数达到最优的可行解称为最优解。在【例 10-3】中，如果能够找到一组能够满足所有约束条件的产量组合，则这个产量组合就是一个可行解；如果这个可行的产量组合能够使得总利润最大，则这个组合便是所求的最优解。

10.3.2.2 线性规划的图解法

【例 10-3】的可行域可用图来描述。该问题有 4 个约束条件

$$
\begin{aligned}
6X+2Y &\leqslant 1\,800 &\quad &\text{(原材料 1 约束)}\\
Y &\leqslant 350 &\quad &\text{(原材料 2 约束)}\\
2X+4Y &\leqslant 1\,600 &\quad &\text{(劳动时间约束)}\\
X \geqslant 0,\ Y &\geqslant 0 &\quad &\text{(非负约束)}
\end{aligned}
$$

图 10-3 给出了满足上述 4 个约束条件的区域。图中，横坐标为 X(产品 A 的产量)。纵坐标为 Y(产品 B 的产量)。约束不等式 $X\geqslant 0$ 表示以 Y 轴(直线 $X=0$)为界的右半平面；约束不等式 $Y\geqslant 0$ 表示以 X 轴(直线 $Y=0$)为界的上半平面；约束不等式 $6X+2Y\leqslant 1\,800$ 表示坐标平面上以直线 $6X+2Y=1\,800$ 为界的左下半平面；约束不等式 $Y\leqslant 350$ 表示坐标平面上以直线 $Y=350$ 为界的下半平面；约束不等式 $2X+4Y\leqslant 1\,600$ 表示坐标平面上以直线 $2X+4Y=1\,600$ 为界的左下半平面。因此，本问题的可行域，即满足所有 4 个约束条件的解的集合，为上述 5 个半平面的交集，也就是位于第Ⅰ象限的凸多边形 $OABCD$(包括边界)。

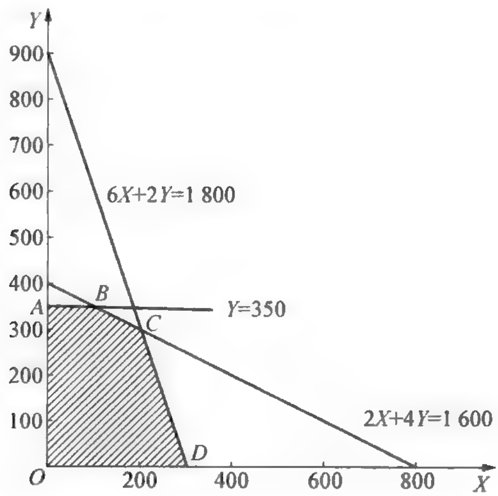


图 10-3 用图解法确定线性规划问题的可行域

本问题的目标是利润最大化，所以应在可行域内选择使得利润达到最大值的解。不妨考虑一下，哪些解可以使利润达到 720 元？要回答这个问题，只需做出等

利润直线 $3X+8Y=720$。该等利润直线上的每一点表示一种产量组合(即一个解),且所有这些点产生的利润均为 720 元。不过,这些点并不都是可行解,只有落在可行域内的那些点才是可行解。同理,作出等利润直线 $3X+8Y=1\,200$,该等利润直线落在可行域内的那些点即为使得利润达到 1 200 元的可行解。如此类推,可以作出一系列等利润直线 $3X+8Y=k$,其中 k 可以取不同的值以表示不同的利润,如图 10-4 所示。

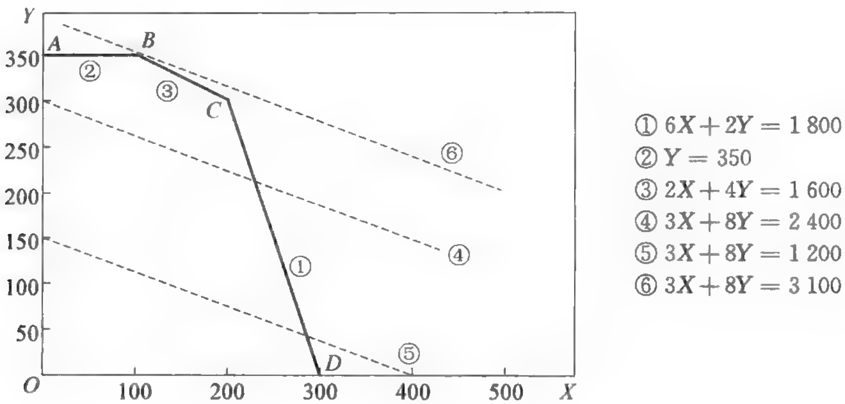


图 10-4 用图解法确定线性规划问题的最优解

不难发现,所有这些等利润直线都相互平行(这是因为它们具有相同的斜率),而且,越远离原点的等利润直线,它所代表的利润越高。因此,最优解应该是在可行域内的最远离原点的那条等利润直线上的点;本题中,既在可行域内,又最远离原点的等利润直线上的点是 B 点,因此 B 点是本问题的最优解,如图 10-4 所示。而 B 点是约束条件直线②(原材料 2 约束)和约束条件直线③(劳动时间约束)的交点,即同时满足下述方程的点:

$$Y=350 \qquad ②$$

$$2X+4Y=1\,600 \qquad ③$$

解上述二元一次方程组,可得最优解为:$X=100$,$Y=350$。相应的最优值为:$3X+8Y=3\times100+8\times350=3\,100$。如果图画得很准确,可以通过直接观察图中 B 点的坐标,得到相同的结果。

图解法适用于求解含有两个决策变量的线性规划问题。现将采用图解法求解线性规划问题的步骤归纳如下:

第一步:在坐标图上做出代表各约束条件的直线;

第二步：确定满足所有约束条件的可行域；

第三步：做出任意一条等利润直线(令利润函数值等于任意一个特定值)；

第四步：朝着使目标函数最优化的方向，平行移动该等利润直线，直到再继续移动就会离开可行域为止。这时，该等利润直线在可行域内的那些点，即为最优解。

10.3.2.3　松弛变量与线性规划模型的标准式

在【例 10－3】中，若将最优解 $X=100$，$Y=350$ 代入约束条件的左边，就可得到三种材料的实际使用量如下：

$$\begin{aligned}
&6X+2Y=1\,300\leqslant 1\,800 \qquad &&(\text{原材料 1 约束}) \quad ①\\
&Y \qquad\qquad\quad \leqslant 350 &&(\text{原材料 2 约束}) \quad ②\\
&2X+4Y \qquad \leqslant 1\,600 &&(\text{劳动时间约束}) \quad ③
\end{aligned}$$

可见原材料 1 有多余，而原材料 2、劳动时间没有多余；约束①称为“非紧”的约束，表示这时资源尚有多余；约束②与③称为“紧”的约束，表示这时资源已全部使用完毕。

若在约束条件左边加上一个变量，即可使得原来的“≤”约束不等式变为等式约束。【例 10－3】中的模型可写为如下的等式形式：

$$\begin{aligned}
\max \quad & F=3X+8Y+0S_1+0S_2+0S_3\\
\text{s.t.} \quad & 6X+2Y+S_1=1\,800\\
& Y+S_2=350\\
& 2X+4Y+S_3=1\,600\\
& X,Y,S_1,S_2,S_3\geqslant 0
\end{aligned}$$

上述等式形式的模型称为标准型线性规划模型。变量 S_1、S_2、S_3 称为松弛变量。松弛变量的值等于“≤”不等式右边的值减去左边的值，它表示可提供的资源与实际消耗的资源之差，即闲置的那部分资源。在本例中，松弛变量的值为

$$S_1=1\,800(\text{约束条件 ① 右边的值})-1\,300(\text{约束条件 ① 右边的值})=500$$
$$S_2=350(\text{约束条件 ② 右边的值})-350(\text{约束条件 ② 右边的值})=0$$
$$S_3=1\,600(\text{约束条件 ③ 右边的值})-1\,600(\text{约束条件 ③ 右边的值})=0$$

10.3.3　最小化问题

前面讨论了线性规划中的最大化问题，本节讨论最小化问题。

10.3.3.1 最小化问题及其线性规划模型

下面讨论一个简单的最小化问题。

【例 10-4】 某港口机械厂成本优化问题。

某港口机械厂生产 A、B 两种港口机械零配件，需要耗费Ⅰ、Ⅱ两种原材料，其价格如表 10-2 所示。

据预测，零配件 B 的需求量不少于 420 件，而零配件 A 由于销路问题，该厂决定，其产量不得超过 600 件。此外，原材料Ⅱ由于库存积压，要求其使用量不得少于 800 吨。

问应使用各种原材料各多少吨，使得在满足要求的前提下总费用最小?

表 10-2 原材料使用情况与价格表

零配件 / 原材料	零配件 A	零配件 B	价格/(元/吨)
原材料Ⅰ	0.40	0.42	45
原材料Ⅱ	0.25	0.15	10

【解】 根据题意可做如下分析:

(1) 决策变量：本问题的决策变量是两种原材料的使用量。可设:

X 为原材料Ⅰ的使用量(吨);

Y 为原材料Ⅱ的使用量(吨)。

(2) 目标函数：本问题的目标函数是总费用最小。总费用可计算如下:

$$\text{总费用} = 45X + 10Y \text{ (元)}$$

(3) 约束条件：本问题共有 4 个约束。第一个约束是零配件 A 的产量约束，第二个约束是零配件 B 的需求约束，第三个约束是原材料Ⅱ的使用量约束，第四个约束是非负约束。由题意，这些约束可表达如下:

$$\begin{aligned} 0.40X + 0.25Y &\leqslant 600 \\ 0.42X + 0.15Y &\geqslant 420 \\ Y &\geqslant 800 \\ X \geqslant 0,\ Y &\geqslant 0 \end{aligned}$$

由上述分析，可建立该最小化问题的线性规划模型如下:

o. b. $\quad \min \quad 45X + 10Y$

s. t. $\quad 0.40X + 0.25Y \leqslant 600 \quad$ (零配件 A 的产量约束)

$\quad 0.42X + 0.15Y \geqslant 420 \quad$ (零配件 B 的需求约束)

$$Y \geqslant 800 \qquad \text{(原材料Ⅱ的使用量约束)}$$
$$X \geqslant 0, Y \geqslant 0 \qquad \text{(非负约束)}$$

10.3.3.2 最小化问题的图解法

图 10-5 给出了满足上述 4 个约束条件的区域。图中，横坐标为 X(零配件 A 的产量)，纵坐标为 Y(零配件 B 的产量)。约束不等式 $X \geqslant 0$ 表示以 Y 轴(直线 $X=0$) 为界的右半平面；约束不等式 $Y \geqslant 0$ 表示以 X 轴(直线 $Y=0$) 为界的上半平面；约束不等式 $0.40X+0.25Y \leqslant 600$ 表示坐标平面上以直线 $0.40X+0.25Y=600$ 为界的左下半平面；约束不等式 $0.42X+0.15Y \geqslant 420$ 表示坐标平面上以直线 $0.42X+0.15Y=420$ 为界的右上半平面；约束不等式 $Y \geqslant 800$ 表示坐标平面上以直线 $Y=800$ 为界的上半平面。因此，本问题的可行域，即满足所有 4 个约束条件的解的集合，为上述 5 个半平面的交集，也就是如图 10-5 所示的位于第Ⅰ象限的凸多边形 ABC(包括边界)。

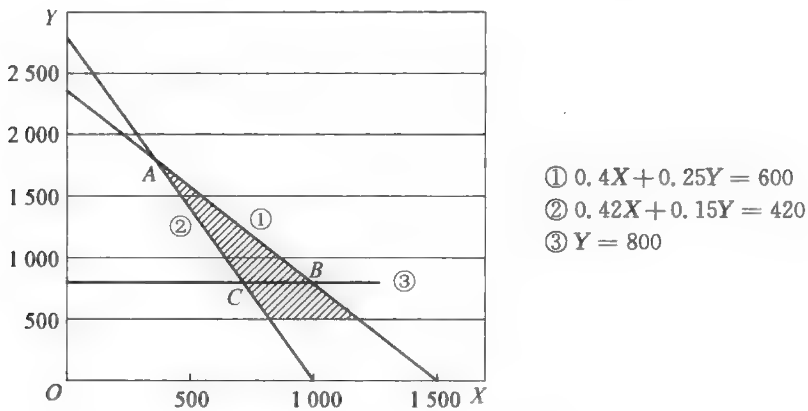


图 10-5 用图解法确定最小化问题的可行域

本问题的目标是费用最小化，所以应在可行域内选择使得费用达到最小值的解。做等费用直线族 $45X+10Y=k$ (k 可取不同的常数)。对于一条等费用直线而言，其上的一个点表示零配件产量的一种组合(即一个解)，且同一条等费用直线上所有各点对应的费用均相等。对于由许多等费用直线组成的等费用直线族来说，越靠近原点的等费用直线对应的费用越小。因此，最优解应是在可行域内的、最接近原点的那条等费用直线上的点。本题中，既在可行城内，又在最接近原点的等费用直线上的点是 A 点，如图 10-6 所示，所以 A 点的坐标就是最优解。而 A 点是约束条件直线①(零配件 A 的产量约束)和约束条件直线②(零配件 B 的需求量约束)的交点，也就是同时满足下述方程的点：

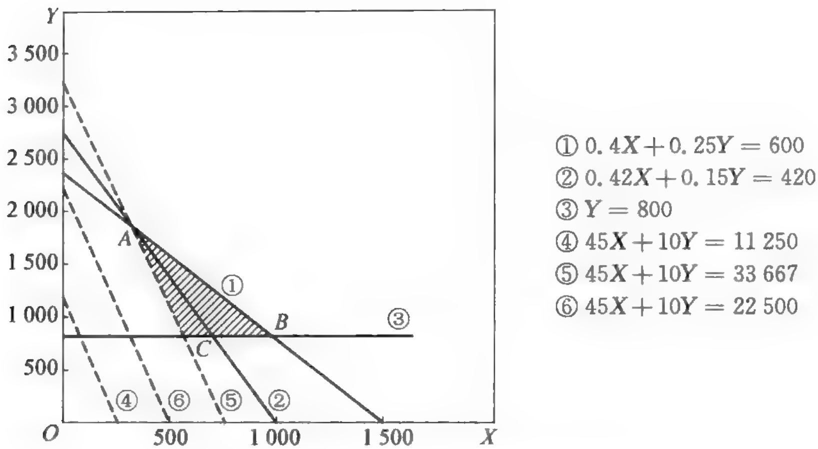


图 10－6　用图解法确定最小化问题的最优解

s. t.
$$0.40X + 0.25Y = 600$$
$$0.42X + 0.15Y = 420$$

解上述二元一次方程组，可得最优解为：$X \approx 333.3$（吨），$Y \approx 1\,866.7$（吨）。相应的最优值为：$45X + 10Y = 45 \times 333.3 + 10 \times 1\,866.7 \approx 33\,666$（元）。该解与图 10－6 中得到的结果相同。

10.3.3.3　剩余变量

在【例 10－4】中，若将最优解 $X = 333.3$、$Y = 1\,866.7$ 代入约束条件的左边，可得：

$$0.40X + 0.25Y = 600 \leqslant 600 \quad \text{（零配件 A 的产量约束）} \quad ①$$
$$0.42X + 0.15Y = 420 \geqslant 420 \quad \text{（零配件 B 的需求约束）} \quad ②$$
$$Y = 1\,867 \geqslant 800 \quad \text{（原材料 Ⅱ 的使用量约束）} \quad ③$$

约束①与②的左边等于右边，称为"紧"的约束；而约束③的左边大于右边，说明原材料Ⅱ的实际使用量不仅能满足所要求的最小使用量，而且有剩余，称约束③为"非紧"约束。

若在约束③的左边减去一个变量，即可使得原来的"≥"约束不等式变为等式约束。同理，可在约束条件②的左边减去一个变量，使原来的"≥"约束不等式变为等式约束；在约束条件①的右边加上一个变量，使原来的"≤"约束不等式变为等式约束。【例 10－4】中的模型则可写为如下的等式形式：

$$
\begin{array}{llllll}
\text{o. b.} & \min & 45X + & 10Y + 0S_1 + 0S_2 + 0S_3 & & \\
\text{s. t.} & & 0.40X + 0.25Y + & S_1 & & = 600 \\
& & 0.42X + 0.15Y & & -S_2 & = 420 \\
& & \quad Y & & -S_3 & = 800 \\
& & X,\ Y,\ S_1,\ S_2,\ S_3 \geqslant 0 & & &
\end{array}
$$

上述等式形式的模型称为标准型线性规划模型。变量 S_2、S_3 称为剩余变量，变量 S_1 称为松弛变量。剩余变量的值等于“$\geqslant$”不等式左边的值减去右边的值，它表示实际完成量与要求完成量之差，即“超额”的或“有剩余”的那部分数量。

在【例 10 - 4】中，约束条件①的松弛变量为

$$S_1 = 600(\text{约束条件右边的值}) - 600(\text{约束条件左边的值}) = 0$$

约束条件②和③的剩余变量分别为

$$S_2 = 420(\text{约束条件左边的值}) - 420(\text{约束条件右边的值}) = 0$$
$$S_3 = 1\,867(\text{约束条件左边的值}) - 800(\text{约束条件右边的值}) = 1\,067$$

10.3.4 线性规划问题的解的讨论

在前面所讨论的例题中，都得到了唯一的最优解。但是，并非所有的线性规划问题都具有唯一解，下面讨论线性规划问题的解可能出现的几种情况。

10.3.4.1 唯一解

线性规划问题具有唯一解是指，该规划问题有且仅有一个既在可行域内，又使目标值达到最优的解。【例 10 - 3】就是一个具有唯一解的规划问题，其数学模型如下：

$$
\begin{array}{llrl}
\text{o. b.} & \max & F = 3X + 8Y & \\
\text{s. t.} & & 6X + 2Y \leqslant 1\,800 & (\text{原材料 1 约束}) \\
& & Y \leqslant 350 & (\text{原材料 2 约束}) \\
& & 2X + 4Y \leqslant 1\,600 & (\text{劳动时间约束}) \\
& & X \geqslant 0,\ Y \geqslant 0 & (\text{非负约束})
\end{array}
$$

上述模型可用图解法求出最优解，如图 10 - 7 所示。

从图 10 - 7 可见，既在可行域 $OABCD$ 中，又使得目标值最大的点只有一个，那就是 B 点。所以 B 点的坐标 $(X = 100,\ Y = 350)$ 是该规划问题唯一的最优解。

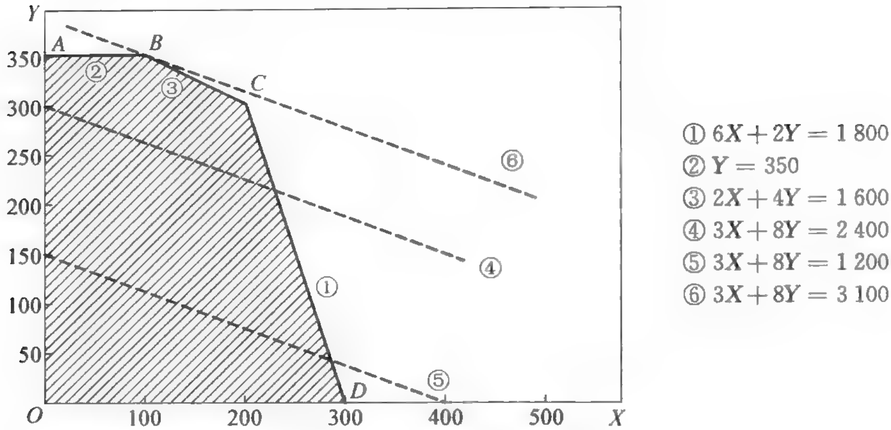


图 10－7　线性规划问题有唯一解的情况

10.3.4.2　无穷多解

线性规划问题具有无穷多解是指，该规划问题有无穷多个既在可行域内、又使目标值达到最优的解。

在【例 10－3】中，设产品 A 的单位产品利润从 3 增加至 4，这时该问题的解将发生变化。用图解法可求出该问题的最优解，如图 10－8 所示。

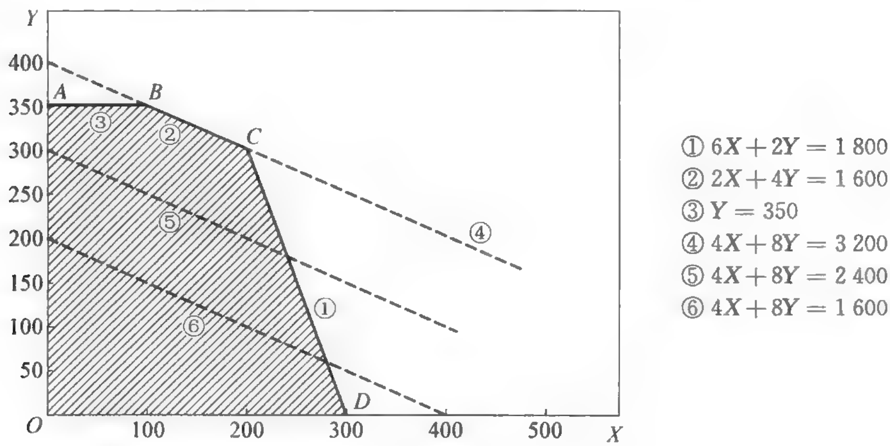


图 10－8　线性规划问题有无穷多个解的情况

从图 10－8 可见，等利润直线族中的直线 $4X+8Y=3\,200$ 与可行域中的边 BC 重叠，这时，线段 BC 上所有的点均为最优解。因此，该规划问题有无穷多个最优解。

10.3.4.3 线性规划问题无可行域的情况

当线性规划问题中的约束条件不能同时满足时，将出现无可行域的情况，这时不存在可行解，即该线性规划问题无解。

在【例 10－3】中，若要求产品 A 的产量不得小于 400，则需再加上一个约束条件 $X \geqslant 400$。从图 10－9 可见，约束条件要求问题的解既在区域 $ABCDO$ 内，又在直线 EF 的右半平面内，显然这是不可能同时满足的，可见这时无可行域。因此，该线性规划问题无解。

有无可行域取决于约束条件，而与目标函数无关。

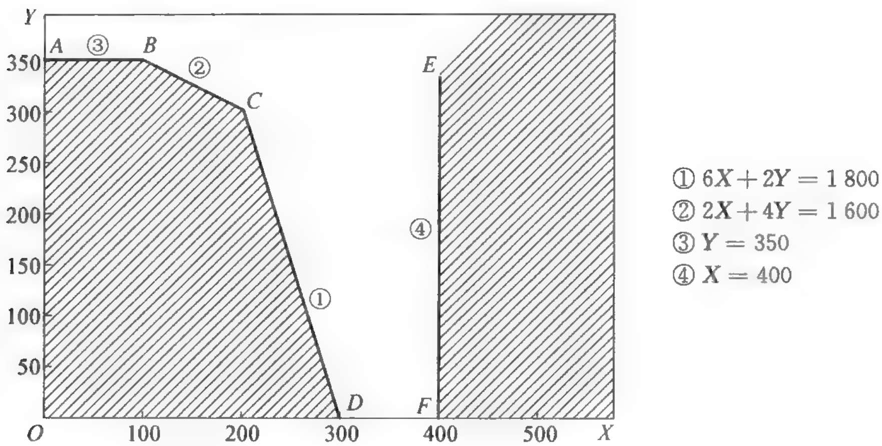


图 10－9 线性规划问题无可行域的情况

10.3.4.4 线性规划问题可行域无界的情况

线性规划问题的可行域无界，是指最大化问题中的目标函数值可以无限增大，或最小化问题中的目标函数值可以无限减小。

在【例 10－3】中，如果没有原材料约束与劳动时间约束，但要求产品 A 与产品 B 的总产量不得少于 350 单位，则该模型变为

$$
\begin{aligned}
&\text{o. b.} \quad \max \quad F = 3X + 8Y \\
&\text{s. t.} \qquad\qquad X + Y \geqslant 350 \quad \text{（产量约束）} \\
&\qquad\qquad\qquad X \geqslant 0,\ Y \geqslant 0 \quad \text{（非负约束）}
\end{aligned}
$$

该问题可用图 10－10 表示。

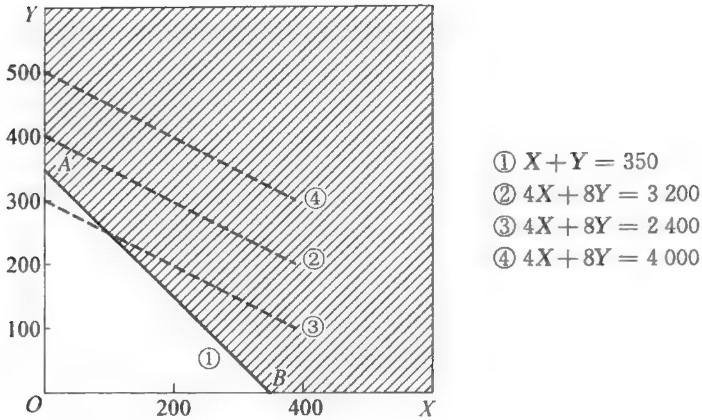


图 10-10 线性规划问题无界的情况

从图 10-10 可见，该问题的可行域是位于直线 AB 右上部分的半平面。在该可行域内，目标函数值(本问题中为利润)可以无限增大，因此该线性规划问题的可行域无界。

10.3.5 线性规划的灵敏度分析和影子价格

本节讨论线性规划问题的灵敏度分析的内容、灵敏度报告的解读与使用以及影子价格的概念等。首先回顾一下【例 10-3】的数学模型。在【例 10-3】中已建立了如下的数学模型：

o. b.	max	$F = 3X + 8Y$		
s. t.		$6X + 2Y \leqslant 1\,800$	(原材料 1 约束)	①
		$Y \leqslant 350$	(原材料 2 约束)	②
		$2X + 4Y \leqslant 1\,600$	(劳动时间约束)	③
		$X \geqslant 0,\ Y \geqslant 0$	(非负约束)	

在上述模型中，决策变量 X、Y 分别表示产品 A 与产品 B 的产量；目标函数是利润最大化，其中，目标函数的系数 3 和 8 分别表示产品 A 与产品 B 每单位产品的利润；约束条件①、②分别是原材料 1 和原材料 2 的供应量约束，约束条件③是劳动时间约束。

经求解，得到上述问题的最优解为：$X = 100$，$Y = 350$。这时，利润达到最大，即得到最优目标值为 3 100 元。

现在假定市场状况或生产工艺发生了变化，使得目标函数中的系数发生了变

化，例如，产品 A 的利润系数从 3(元/单位产品)增至 3.5(元/单位产品)，那么，已求得的最优解、最优目标值会发生变化吗，目标函数的系数在什么范围内变化，才不会影响最优解？另外，如果原材料或劳动时间的供应量增加，最大利润将会如何变化？这些问题在实际生产管理中是十分重要的，它们也是灵敏度分析所要回答的问题。

10.3.5.1 灵敏度分析的内容

灵敏度分析是研究当目标函数中的系数发生变化以及当约束条件右边的值发生变化时，原有的最优解、最优目标值受到的影响。

1）目标函数中的系数的变化对最优解与最优目标值的影响

当目标函数中的系数变化时，等利润直线变得陡峭或平坦，它与可行域的交点也可能随之变化。目标函数中的系数改变足够大时，可使最优解发生变化。图 10-11 描绘了【例 10-3】中目标函数中的系数变化时等利润直线的变化。当产品 A 的单位产品利润变化时，该等利润直线的斜率将发生变化。若等利润直线在图中所示的两条虚线 AE 和 BF 之间的范围内变化，则 B 点仍然是既在可行域上，又离原点最远的顶点，因此，这时最优解保持不变；若等利润直线变得足够陡峭(或足够平坦)，超出了两条虚线之间的范围，则该等利润直线将与可行域相交于另一顶点 C 点(或 A 点)，这时最优解将从顶点 B 点变为另一个顶点 C 点(或 A 点)。

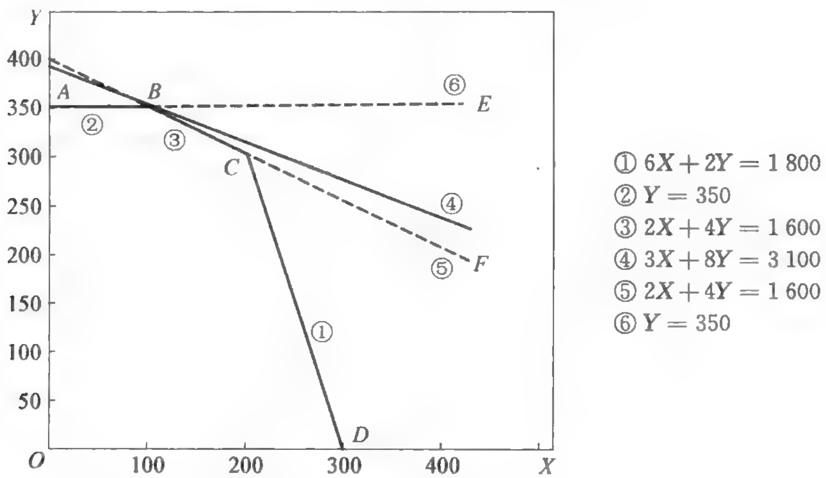


图 10-11 目标函数中的系数变化对最优解的影响

可见，当目标函数中的系数发生变化时，若变化量在某个范围内，则最优解不变；若变化足够大，则最优解将发生变化。而当最优解发生变化时，通常最优目标值也将随之发生变化。

2）约束条件右边的变化对最优解与目标值的影响

当约束条件右边变化时，相应的表示约束的直线将平行移动，可行域将发生变化。当该移动足够大时，最优解、目标值也可能随之变化。图 10－12 描绘了【例 10－3】中约束条件③（劳动时间约束）右边发生变化时可行域的变化。当可提供的劳动时间减少时，表示劳动间约束条件的直线 BC 移动至图中虚线 $B'C'$ 所示的位置，可行域亦随之变化，从多边形 $OABCD$ 变为多边形 $OAB'C'D$。这时，最优解与目标值均将发生变化。但是，如果约束条件①（原材料 1 约束）的右边发生变化，而且变化不太大，则可行域的变化不会影响最优解与目标值，该约束条件是“非紧”的。当然，如果变化很大，以至使该约束条件成为“紧”的，这时，最优解与最优目标值均可能发生变化。

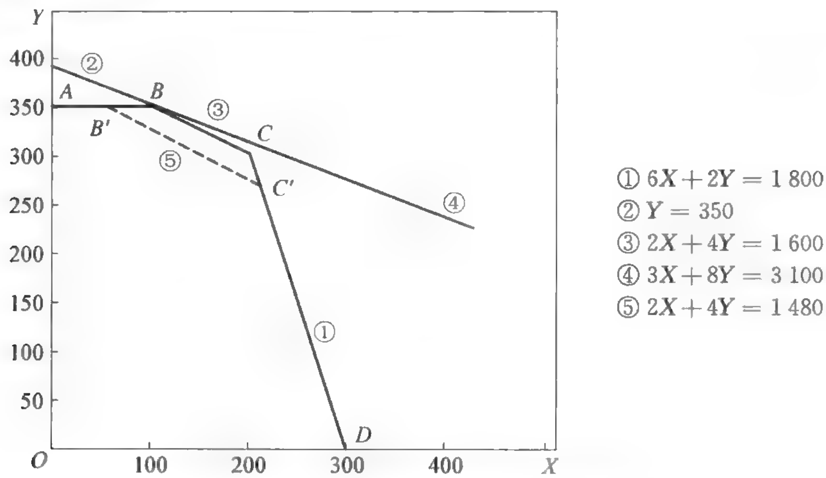


图 10－12　约束条件右边变化对最优解的影响

可见，当约束条件右边发生变化时，最优解与最优目标值可能会发生变化。综上所述，灵敏度分析的主要内容包括：

（1）目标函数中的系数变化时，表示目标函数的直线族变得陡峭或平坦，它与可行域的交点也可能随之变化。灵敏度分析是研究目标函数中的系数变化对最优解与目标值的影响，以及目标函数中的系数改变多少，方可使最优解发生变化。

（2）约束条件右边变化时，相应的表示约束条件的直线将平行移动，可行域发生变化，最优解与最优目标值也可能随之变化。灵敏度分析是研究约束条件右边变化对目标值或最优解的影响状况。

10.3.5.2　**敏感性报告**

灵敏度分析所要解决的问题可通过数学方法进行分析，例如可用数学公式计

算目标函数中的系数或约束条件右边变化对最优解与最优目标值的影响。不过，这种计算一般比较复杂。本节主要介绍如何解读 Excel 中“规划求解”模块生成的敏感性报告。关于灵敏度分析的求解方法和过程，不是本书的主要内容，请参照有关运筹学书籍。

1) 敏感性报告中各项指标的含义

仍以【例 10－3】为例，运用 Excel 中的“规划求解”功能可得到敏感性报告，如表 10－3 所示。下面介绍该敏感性报告中各项指标的含义。

表 10－3 敏感性报告

Microsoft Excel 12.0 敏感性报告
工作表 [港口机械生产优化问题.xlsx]最大化问题
报告的建立：2012-11-11 15:53:22

可变单元格

单元格	名字	终值	递减成本	目标式系数	允许的增量	允许的减量
$B$13	产量 机械A	100	0	3	1	3
$C$13	产量 机械B	350	0	8	1E+30	2

约束

单元格	名字	终值	阴影价格	约束限制值	允许的增量	允许的减量
$B$18	原材料1 使用量（左边）	1300	0	1800	1E+30	500
$B$19	原材料2 使用量（左边）	350	2	350	50	50
$B$20	劳动时间 使用量（左边）	1600	1.5	1600	166.6666667	200

敏感性报告由两部分组成。位于报告上部的表格(单元格 A6：I10)是关于目标函数中的系数变化对最优解产生的影响；位于报告下部的表格(单元格 A12：I17)是关于约束条件右边变化对目标值的影响。

位于敏感性报告上部的表格反映目标函数中系数变化对最优解的影响。表格中的前三列是关于问题中决策变量的信息，其中，“单元格”是指决策变量所在单元格的地址，“名字”是这些决策变量的名称，“终值”是决策变量的终值，即最优解。在本题中，两个决策变量：产品 A 的产量与产品 B 的产量，它们在 Spreadsheet 上的地址分别是 $B$13 和 $C$13，其最优解分别为 100 和 350 单位。第四列是“递减成本”，它的绝对值表示目标函数中决策变量的系数必须改进多少，才能得到该决策变量的正数解。这里的“改进”，在最大化问题中是指增加，在最小化问题中则是指减少，本题中，两个决策变量均已得到正数解，所以它们的递减成本均为零。第五列“目标式系数”是指目标函数中的系数，它是题目中的已知条件。本题中，目

标函数中两个决策变量的系数分别为 3 和 8。第六列与第七列分别是“允许的增量”和“允许的减量”，它们表示目标函数中的系数在允许的增量与减量范围内变化时，最优解不变。本题中第一个决策变量(产品 A 的产量)的目标系数为 3，允许的增量为 1，允许的减量为 3，因此，该目标系数在[3－3, 3＋1]，也就是[0, 4]范围内变化时，该问题的最优解不变。同理，第二个决策变量(产品 B 的产量)的目标系数为 8，允许的增量为 1×10^{30}，允许的减量为 2，因此，该目标系数在[6, ＋∞]范围内变化时，该问题的最优解不变。应注意，这里给出的决策变量的“允许变化范围”是指其他条件不变，仅在该决策变量变化时的允许变化范围。

位于敏感性报告下部的表格反映约束条件右边变化对目标值的影响。表格中的前三列是关于约束条件左边的信息。其中，“单元格”是指约束条件左边所在单元格的地址，“名字”是约束条件左边的名称，“终值”是约束条件左边的终值。在本题中，有三个约束条件，它们的左边分别是原材料 1 使用量，原材料 2 使用量和劳动时间使用量，它们在 Spreadsheet 上的地址分别是 \$B\$18，\$B\$19 和 \$B\$20，其终值分别为 1 300，350 和 1 600。第四列为“阴影价格”，即影子价格，关于影子价格问题将在稍后进行讨论。第五列为“约束限制值”，指约束条件右边的值，通常是题目中给出的已知条件。本题中，三个约束条件右边的值分别表示原材料 1、原材料 2 与劳动时间的供应量，它们分别为 1 800，350 与 1 600。第六列与第七列是“允许的增量”和“允许的减量”，它们表示约束条件右边在允许的增量与减量范围内变化时，影子价格不变。例如本题中，第一个约束条件右边的值为 1 800，允许的增量为 1×10^{30}，允许的减量为 500，因此，该约束条件右边在[1 800－500, 1 800＋∞]，也就是[1 300, ＋∞]范围内变化时，原材料 1 的影子价格不变。应注意，这里给出的某约束条件右边的“允许变化范围”是指其他条件不变，仅在该约束条件右边变化时的允许变化范围。同理，第二个约束条件右边在[350－50, 350＋50]，即[300, 400]范围内变化时，原材料 2 的影子价格不变；第三个约束条件右边在[1 600－200, 1 600＋166.7]，即[1 400, 1 766.7]范围内变化时，劳动时间的影子价格不变。

2) 影子价格

在敏感性报告下部的表格中，第四列是影子价格，这是一个十分重要的概念。影子价格是指约束条件右边增加(或减少)一个单位，目标值增加(或减少)的数量。

在【例 10－3】中有三个资源约束，每种资源的影子价格是该种资源供应量增加(或减少)一个单位时，总利润增加(或减少)的数量。例如，从敏感性报告可知，第一个约束条件(原材料 1 供应量约束)的影子价格为 0，这说明在允许的范围[1 300, ＋∞]内，再增加(或减少)一个单位的原材料 1 供应量，总利润不变。第二个约束条件(原材料 2 供应量约束)的影子价格为 2，说明在允许的范围[300, 400]

内，再增加(或减少)一个单位的原材料 2 供应量，总利润将增加(或减少)2 元。第三个约束条件(劳动时间供应量约束)的影子价格为 1.5，说明在允许的范围[1 400, 1 766.7]内，再增加(或减少)一个单位的劳动时间供应量，总利润将增加(或减少)1.5 元。

3) 使用敏感性报告进行灵敏度分析

下面采用敏感性报告对【例 10-3】进行灵敏度分析，并回答本节开始时提出的问题。

(1) 若产品 A 的利润系数从 3(元/单位产品)增至 3.5(元/单位产品)，那么已求得的最优解、最优目标值会变化吗？该系数在什么范围内变化，才不会影响最优解？

由表 10-3 所示的敏感性报告上部的表格可知，产品 A 的系数在允许的变化范围[3—3, 3+1]，即[0, 4]区间变化时，不会影响最优解。现在，产品 A 的利润系数增至 3.5，是在允许的变化范围内，所以最优解不变，仍然是 $X=100$，$Y=350$。

应注意的是，这时最优目标值(即最大利润)将发生变化。原来的最大利润为

$$3X+8Y=3\times 100+8\times 350=3\,100(\text{元})$$

$$\text{变化后的最大利润}=3\,100+(3.5-3)\times 100=3\,150(\text{元})$$

(2) 若原材料 2 的供应量增加 30 kg，最大利润将为多少？

由表 10-3 所示的敏感性报告下部的表格可知，当原材料 2 的约束条件右边在允许变化的范围[350—50, 350+50]，即[300, 400]区间变化时，原材料 2 的影子价格不变。现在，原材料 2 的供应量增加了 30 kg，变为 380 kg，是在允许增加的范围内，所以，其影子价格不变，仍然等于 2。这就是说，原材料 2 的供应量每增加 1 kg 将使最大利润增加 2 元。当原材料 2 的供应量增加 30 kg 时，最大利润将增加 $2\times 30=60$(元)，最大利润 $=3\,100+60=3\,160$(元)。

10.3.5.3 例题

【例 10-5】 港口机械厂生产计划优化问题。

某港口机械厂生产 4 种小型工具，由于该 4 种工具具有不同的大小、形状和功能，所以它们所需要的主要原料(A 和 B)、制作时间、最大销售量与利润均不相同。

该厂每天可提供的原材料 A、B 与工人劳动时间分别为 600 单位、1 000 单位与 400 小时，如表 10-4 所示。问：

(1) 应如何安排该 4 种工具的日产量，使得该厂的日利润最大？

(2) 该厂是否愿意付出 10 元的加班费，让某工人加班 1 小时？

(3) 如果可提供的工人劳动时间变为 398 小时,该厂的日利润将有何变化?

(4) 该厂应优先考虑购买何种资源?

(5) 若因市场变化,第一种工具的单位利润从 60 元下降到 55 元,问该厂的生产计划及日利润将如何变化?

表 10-4 某港口机械厂生产基本数据

工具类型	劳动时间 /(小时/件)	原材料 1 /(单位/件)	原材料 2 /(单位/件)	单位产品利润 /(元/件)	最大销售量 /件
1	2	4	6	60	100
2	1	2	2	20	200
3	3	1	1	40	50
4	2	2	2	30	100
可提供量	400(小时)	600(单位)	1 000(单位)		

【解】 根据题意,本问题的决策变量是 4 种工具的日产量,目标函数是日利润最大化,而约束条件则是资源(原材料 A、B 和劳动时间)约束、需求量约束与非负约束。因此,本问题可用下列线性规划模型描述。

$$
\begin{array}{lll}
\text{o. b.} \quad \max & 60X_1 + 20X_2 + 40X_3 + 30X_4 & \\
\text{s. t.} & 4X_1 + 2X_2 + X_3 + 2X_4 \leqslant 600 & \text{(原材料 1 约束)} \\
& 6X_1 + 2X_2 + X_3 + 2X_4 \leqslant 1\,000 & \text{(原材料 2 约束)} \\
& 2X_1 + X_2 + 3X_3 + 2X_4 \leqslant 400 & \text{(劳动时间约束)} \\
& X_1 \leqslant 100 & \text{(工具 1 需求量约束)} \\
& X_2 \leqslant 200 & \text{(工具 2 需求量约束)} \\
& X_3 \leqslant 50 & \text{(工具 3 需求量约束)} \\
& X_4 \leqslant 100 & \text{(工具 4 需求量约束)} \\
& X_1,\ X_2,\ X_3,\ X_4 \geqslant 0 &
\end{array}
$$

式中,X_1, X_2, X_3, X_4 分别为 4 种工具的日产量。

用 Excel 中的"规划求解"功能可以求出上述线性规划问题的最优解与最优值。"规划求解"的具体使用方法在本书中不做详细描述。运算结果可如表 10-5 所示。

表 10－5　港口机械厂生产优化问题的运算结果

Microsoft Excel 12.0 运算结果报告
工作表 [家具优化问题.xlsx]Sheet1
报告的建立：2012-12-24 17:13:34

目标单元格（最大值）

单元格	名字	初值	终值
$H$9	日利润（元/件） 使用量	9200	9200

可变单元格

单元格	名字	初值	终值
$C$13	日产量（件）	100	100
$D$13	日产量（件）	80	80
$E$13	日产量（件）	40	40
$F$13	日产量（件）	0	0

约束

单元格	名字	单元格值	公式	状态	型数值
$H$5	劳动时间（小时/件） 使用量	400	$H$5<=$J$5	到达限制值	0
$H$6	木材（单位/件） 使用量	600	$H$6<=$J$6	到达限制值	0
$H$7	玻璃（单位/件） 使用量	800	$H$7<=$J$7	未到限制值	200
$C$13	日产量（件）	100	$C$13<=$C$15	到达限制值	0
$D$13	日产量（件）	80	$D$13<=$D$15	未到限制值	120
$E$13	日产量（件）	40	$E$13<=$E$15	未到限制值	10
$F$13	日产量（件）	0	$F$13<=$F$15	未到限制值	100
$C$13	日产量（件）	100	$C$13>=0	未到限制值	100
$D$13	日产量（件）	80	$D$13>=0	未到限制值	80
$E$13	日产量（件）	40	$E$13>=0	未到限制值	40
$F$13	日产量（件）	0	$F$13>=0	到达限制值	0

根据模型运行结果可做出如下分析：

(1) 由模型的解可知，港口机械厂 4 种工具的最优日产量分别为 100 件、80 件、40 件和 0 件，这时该厂的日利润最大，为 9 200 元。

本问题的敏感性报告如表 10－6 所示。由上述敏感性报告可进行灵敏度分析，并回答题目中的问题(2)～(5)。

表 10－6　港口机械厂生产优化问题的敏感性报告

Microsoft Excel 11.0 敏感性报告
可变单元格

单元格	名　字	终值	递减成本	目标式系数	允许的增量	允许的减量
$B$15	日产量/件	100	20	60	1×10^{30}	20
$C$15	日产量/件	80	0	20	10	2.5

续 表

单元格	名 字	终值	递减成本	目标式系数	允许的增量	允许的减量
$D$15	日产量/件	40	0	40	20	5.0
$E$15	日产量/件	0	−2.0	30	2.0	1×10^{30}
约束						
$G$6	劳动时间/(小时/件)	400	12	400	25	100
$G$7	原材料 A/(单位/件)	600	4	600	200	50
$G$8	原材料 B/(单位/件)	800	0	1 000	1×10^{30}	200

(2) 由敏感性报告表 10 - 6 可知，劳动时间的影子价格为 12 元，即在劳动时间的增量不超过 25 小时的条件下，每增加 1 小时劳动时间，该厂的利润(目标值)将增加 12 元。因此，付给某工人 10 元以增加 1 小时劳动时间是值得的，可多获利 $12-10=2$(元)。

(3) 当可提供的劳动时间从 400 小时减少为 398 小时时，该减少量在允许的减量(100 小时)内，所以劳动时间的影子价格不变，仍为 12 元。因此，该厂的利润变为：$9\,200+12\times(398-400)=9\,176$(元)。

(4) 由敏感性报告可见，劳动时间与原材料 A 这两种资源的使用量等于可提供量，所以它们的约束条件为“紧”的，即无余量的，而原材料 B 的使用量为 800，可提供量为 1 000，所以原材料 B 的约束条件是“非紧”的，即有余量的。因此，应优先考虑购买劳动时间与原材料 A 这两种资源。

(5) 由敏感性报告可知，工具 1 的目标系数(即单位利润)允许的减量为 20，即当工具 1 的单位利润减少量不超过 20 元时，最优解不变。因此，若工具 1 的单位利润从 60 元下降到 55 元，下降量为 5 元，该下降量在允许的减量范围内，这时，最优解不变。4 种工具的最优日产量仍分别为 100 件、80 件、40 件和 0 件。最优值变为 $9\,200+(55-60)\times100=8\,700$(元)。

思考与练习

1. 某人持有现金 5 万元，正准备到股票市场购买某种股票。假定他对 6 家公司 (A_i, $i=1, 2, \cdots, 6$) 发行的股票感兴趣，但只想购买其中一种股票。这 6 家公司发行的股票每年每股发放股利 (D_i, $i=1, 2, \cdots, 6$)，如题表 10 - 1 所示：

题表 10 - 1

公司 A_i	每年每股股利 D_i(单位：元)
A_1	180
A_2	184
A_3	182
A_4	183
A_5	185
A_6	181

若购买这 6 家公司发行的股票，除每年每股股利外，其他情况大致相同，那么该人应购买哪家公司发行的股票?

2. 某外资企业在原有设备条件下生产某种产品，产品销售单价为 25 元，单位产品的可变成本为 12 元，年固定成本费用为 5.4 万元。如果更新设备，年固定成本将提高 6.24 万元。但由于先进设备的采用，单位可变成本可降为 10 元。若年产量为 4 300 件，为使该企业获得更高的利润，设备是否应更新?

3. Quality 空调制造公司生产 3 种空调：经济型、标准型和高级型。每种空调的利润分别是 63 美元、95 美元和 135 美元。空调的安全生产要求如下：

题表 10 - 2 生产要素表

	风扇马达数	制冷盘数	生产时间(小时)
经济型	1	1	8
标准型	1	2	12
高级型	1	4	14

对于即将到来的生产期，公司拥有的资源是风扇马达 200 个、制冷盘 320 个和生产时间 2 400 小时。那么如何安排生产才能使总利润最大? 请根据题表 10 - 3、题表 10 - 4 回答：

(1) 根据上述描述写出线性规划模型

(2) 最优解是什么? 目标函数值是多少?

(3) 哪些资源被完全利用，为什么?

(4) 哪些资源有剩余? 剩余多少?

(5) 如果标准型产品的利润增加到 100 美元/台，最优解将如何变化?

(6) 如果经济型产品的利润增加到 80 美元/台，最优解如何变化?

(7) 如果制冷盘总量减少到 300 个，则最大的总利润会有什么变化?

题表 10-3 运算结果报告

目标单元格（最小值）

单元格	名 字	初 值	终 值
$B$21	Total profit	$0	$16 440

可变单元格

单元格	名 字	初 值	终 值
$C$13	经济型	0	80
$D$13	标准型	0	120
$E$13	高级型	0	0

约束

单元格	名 字	单元格值	公 式	状 态	型数值
$C$16	风扇马达	200	$C$16>=$C$18	到达限制值	0
$D$16	制冷盘	320	$D$16>=$D$18	到达限制值	0
$E$16	生产时间	2 080	$E$16>=$E$18	未到达限制值	320

题表 10-4 敏感性报告

可变单元格

单元格	名 字	终值	递减成本	目标式系数	允许的增量	允许的减量
$C$13	经济型	80	0	63	12	15.5
$D$13	标准型	120	0	95	31	8
$E$13	高级型	0	7	135	24	无下限

约束

单元格	名 字	终值	阴影价格	约束限制值	允许的增量	允许的减量
$C$16	风扇马达	200	0	200	80	40
$D$16	制冷盘	320	8	320	80	120
$E$16	生产时间	2 080	15	2 400	无上限	320

11　非确定型决策

非确定型决策是指决策者对未来虽然有一定程度的了解，知道可能出现的各种自然状态，但又无法确定各种自然状态可能发生的概率的情况下的决策。这种决策，由于有关因素难以计算，因此完全取决于决策者的经验、判断和估计，其决策准则带有某种程度的主观随意性。

11.1　最大最小决策准则

11.1.1　最大最小决策准则的概念

"最大最小"决策准则也称为"小中取大"决策准则或称为悲观决策准则。这种决策准则的客观依据是决策的系统功能欠佳，形式对决策者不利，所以决策者没有理由希望获得最理想的结果。而面对这种情况，决策者必须从每一方案的最坏处着眼，从每个方案的最坏结果中选择一个最佳值，即在所有最不利的收益中，选取一个收益最大的方案作为决策方案。这种决策方法是十分保守的。

设有一非确定性决策，备选方案为 $d_i(i=1, 2, \cdots, m)$，自然状态有 n 种(其出现概率未知)，损益值为 $L_{ij}(i=1, 2, \cdots, m; j=1, 2, \cdots, n)$，若 $f(d_i)$表示采取行动方案 d_i 时的最小收益，即

$$f(d_i)=\min(L_{i1}, L_{i2}, \cdots, L_{in})\quad (i=1, 2, \cdots, m)$$

则满足

$$f(d_*)=\max[f(d_1), f(d_2), \cdots, f(d_m)]$$

的方案 d_*，就是"最大最小"决策的最优方案。

若决策矩阵为损失矩阵，则在采取"最大最小"的方法时，$f(d_i)$表示采取行动方案 d_i 的最大损失值，即

$$f(d_i)=\max(L_{i1}, L_{i2}, \cdots, L_{in})\quad (i=1, 2, \cdots, m)$$

则满足

$$f(d_*)=\min[f(d_1), f(d_2), \cdots, f(d_m)]$$

的方案 d_*，就是"最大最小"决策的最优方案。

11.1.2 最大最小决策方法的应用

下面我们举两个例子分别说明收益和损失这两种不同情况下决策方法的应用过程。

【例 11-1】 某港口集团公司决定在自己下属码头公司范围内进行新码头管理方式试点，以提高某码头的生产效率。可供选择的行动方案有以下 4 种：d_1，引进新的港口管理信息系统；d_2，购买新型岸边装卸机械；d_3，购买新型堆场作业机械；d_4，扩大码头堆场面积。由于资金所限，试点时只能实施以上 4 种方式中的一种。由于在下属码头公司范围内有 A、B、C 三个不同试点(该三个试点装卸的货种一致)，初步估计出每种方案对于不同公司有其不同的收益值，如表 11-1 所示。试做出决策，使所获收益为最大。

表 11-1 各种不同方案的收益值

自然状态 / 收益值/(元/吨) / 方案	A 公司 θ_1	B 公司 θ_2	C 公司 θ_3
新系统 d_1	105	60	92
新型岸边机械 d_2	55	80	120
新型堆场机械 d_3	70	80	100
扩大堆场面积 d_4	80	65	130

在本例中，A、B、C 三个公司可以看作是该决策系统中三种不同的自然状态。由于没有确定在哪一个公司试点，所以，各公司的生产产量在总公司中所占的比例即概率是不可能知道的。因此，这是一个不确定型决策问题。同时，决策者认为客观形势不利，因此，采用保守的决策方法，即采用"最大最小"的决策方法进行分析。

各方案的最小收益值为

$$f(d_1) = \min(105, 60, 92) = 60$$
$$f(d_2) = \min(55, 80, 120) = 55$$
$$f(d_3) = \min(70, 80, 100) = 70$$
$$f(d_4) = \min(80, 65, 130) = 65$$

这些最小收益值中的最大者为

$$f(d_*) = \max(60, 55, 70, 65) = 70$$

所以，最佳决策为 $d_* = d_3$，即采用新型堆场机械方案 d_3，这无论对于选哪一个公司做试点，都可望获得较大的收益。

上例的决策过程也可以通过收益矩阵表来进行，如表 11-2 所示。

表 11-2 收益矩阵决策表

收益值/(元/吨) 自然状态 方案	A公司 θ_1	B公司 θ_2	C公司 θ_3	$\min\limits_{\theta_j}[L_{ij}]$
新系统 d_1	105	60	92	60
新型岸边机械 d_2	55	80	120	55
新型堆场机械 d_3	70	80	100	70
扩大堆场面积 d_4	80	65	130	65
决　策	$\max\limits_{d_i}\{\min\limits_{\theta_j}[L_{ij}]\}$			70

若决策矩阵是损失矩阵，则应采用“最小最大”决策方法。

【例 11-2】 某港口机械零配件厂拟对其生产的某种机器零配件是否明年改型以及怎样改型做出决策，有三个方案可供选择：方案 d_1，机芯、机壳同时改型；方案 d_2，机芯改型，机壳不改型；方案 d_3，机壳改型，机芯不改型。改型后的机器可能遇到三种自然状态：高需求 θ_1、中需求 θ_2 和低需求 θ_3，其费用损失矩阵如表 11-3 所示。试问应该怎样决策？

表 11-3 费用损失矩阵表

费用/万元 自然状态 方案	高需求 θ_1	中需求 θ_2	低需求 θ_3
d_1	0	16.5	21.5
d_2	22.5	0	13.5
d_3	27.5	17.5	0

由于竞争厂家较多，因此，决策者对该零配件的销售前景抱悲观态度。决策过程如表 11-4 可知，应选择方案 d_1，它在最不利情况下的费用损失为最小。

表 11-4 费用损失矩阵决策过程

费用/万元 自然状态 方案	高需求 θ_1	中需求 θ_2	低需求 θ_3	$\max\limits_{\theta_j}[L_{ij}]$
d_1	0	16.5	21.5	21.5
d_2	22.5	0	13.5	22.5
d_3	27.5	17.5	0	27.5
决　策	$\min\limits_{d_i}\max\limits_{\theta_j}[L_{ij}]$			21.5

11.2 最大最大决策准则

11.2.1 最大最大决策准则的概念及步骤

“最大最大”决策准则也称为乐观决策准则，或称为“好中求好”决策准则。这种决策准则就是充分考虑可能出现的最大利益，在各最大利益中选取最大者，将其对应的方案作为最优方案。这种决策准则的客观基础就是所谓的天时、地利和人和，决策者感到前途乐观，有信心取得每一决策方案的最佳结果。

“最大最大”决策方法的一般步骤为：

(1) 确定各种可行方案。

(2) 确定决策问题将面临的各种自然状态。

(3) 将各种方案在各种自然状态下的损益值列于决策矩阵表中。

设某一决策问题有 m 个行动方案 $d_1, d_2, \cdots, d_m$，n 个自然状态 $\theta_1, \theta_2, \cdots, \theta_n$，损益值 $L_{ij}(i=1, 2, \cdots, m;\ j=1, 2, \cdots, n)$，则“最大最大”的决策矩阵如表 11-5 所示。

表 11-5 “最大最大”决策矩阵表

自然状态 / 损益值 / 行动方案	θ_1 θ_2 $\cdots$ θ_n	$\max\limits_{\theta_j}[L_{ij}]$
d_1	L_{11} L_{12} $\cdots$ L_{1n}	
d_1	L_{21} L_{22} $\cdots$ L_{2n}	
$\vdots$	……	
d_m	L_{m1} L_{m2} $\cdots$ L_{mn}	
决　策	$\max\limits_{d_i}\{\max\limits_{\theta_j}[L_{ij}]\}$	

(4) 求每一方案在各自然状态下的最大损益值：

$$\max\{L_{11}, L_{12}, \cdots, L_{1n}\}$$
$$\max\{L_{21}, L_{22}, \cdots, L_{2n}\}$$
$$\cdots\cdots$$
$$\max\{L_{m1}, L_{m2}, \cdots, L_{mn}\}$$

将其填写在决策矩阵表的最右一列。

(5) 取 $\max\limits_{\theta_j}[L_{ij}]$ 中的最大值 $\max\limits_{d_i}\{\max\limits_{\theta_j}[L_{ij}]\}$，所对应的方案 d_i 为最佳决策

方案。如果表 11－5 不是收益矩阵，而是损失矩阵，则应采取“最小最小”决策准则，取 $\min\limits_{\theta_j}[L_{ij}]$ 中的最小值 $\min\limits_{d_i}\{\min\limits_{\theta_j}[L_{ij}]\}$，所对应的方案 d_i 就是最佳决策方案。

11.2.2 最大最大决策方法的应用

【例 11－3】 例如，某集装箱制造厂根据自己的生产能力制订了以下三个可供选择的行动方案：d_1，生产 20′、40′和 40′HG 集装箱各 1.0 万个；d_2，生产 20′集装箱 1.2 万个，40′集装箱 1.0 万个，40′HG 集装箱 0.7 万个；d_3，生产 20′集装箱 0.8 万个，40′集装箱 0.9 万个，40′HG 集装箱 1.4 万个。各方案的优劣主要取决于市场销售情况，而市场销售情况与各航运物流企业对各种类型集装箱的购买能力有关。根据预测，市场情况可能会有以下状态：θ_1，40′HG 集装箱市场销售量较少；θ_2，40′HG 集装箱销售量大，40′集装箱销售量较大；θ_3，40′HG 集装箱销售量较大，20′和 40′集装箱销售量基本持平。制造厂决策者认为，该厂生产的集装箱有一定的市场信誉和广告宣传优势，又有一定的后备资金，因此前途乐观。其决策过程如下：

(1) 计算各行动方案的利润值，各种状态下的生产成本、销售价、销售额和利润情况如表 11－6 所示。

表 11－6 各行动方案利润值 单位：万个、百万元

方案	品种	生产量	销售价	成本	销售量			销售额			利润		
					θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
方案 d_1	20′	1	4.3	1.5	1.0	0.9	0.8	4.30	3.87	3.44	2.80	2.37	1.94
	40′	1	4.5	1.5	1.0	1.0	0.9	4.50	4.50	4.05	3.00	3.00	2.55
	40′HG	1	3.7	1.0	0.7	1.0	1.0	2.59	3.70	3.70	1.59	2.70	2.70
总计		—		—	4.0		—	11.39	12.07	11.19	7.39	8.07	7.19
方案 d_2	20′	1.2	4.3	1.8	1.2	0.9	0.8	5.16	3.87	3.44	3.36	2.07	1.61
	40′	1.0	4.5	1.5	1.0	1.0	0.9	4.50	4.50	4.05	3.00	3.00	2.55
	40′HG	0.7	3.7	0.7	0.7	0.7	0.7	2.59	2.59	2.59	1.89	1.89	1.89
总计		—	—	4.0	—	—	-	12.25	10.96	10.08	8.25	6.96	6.08
方案 d_3	20′	0.8	4.3	1.2	1.2	0.8	0.8	3.44	3.44	3.44	2.24	2.24	2.24
	40′	0.9	4.5	1.35	1.35	0.9	0.9	4.05	4.05	4.05	2.70	2.70	2.70
	40′HG	1.4	3.7	1.4	1.4	0.7	1.0	2.59	5.18	3.70	1.19	3.78	2.30
总计		—	—	3.96	—	—	—	10.08	12.67	11.19	6.13	8.72	7.24

(2) 由表 11 - 6 可列出决策矩阵表，如表 11 - 7 所示。

表 11 - 7 决 策 矩 阵 表

利润/万元 状态 方案	θ_1	θ_2	θ_3	$\max\limits_{\theta_j}[L_{ij}]$
d_1	7.39	8.07	7.19	8.07
d_2	8.25	6.96	6.08	8.25
d_3	6.13	8.72	7.24	8.72
决　策	$\max\limits_{d_i}\{\max\limits_{\theta_j}[L_{ij}]\}$			8.72

由表 11 - 7 可知，若选择第一方案 d_1，则在第二种状态 θ_2 时可获得最大利润 8.07 百万元；若选择第二方案 d_2，则在第一种状态 θ_1 时可获得最大利润 8.25 百万元；若选择第三方案 d_3，则在第二种状态 θ_2 时可获得最大利润 8.72 百万元。决策者希望获得最大利润，故选择第三方案。虽然第三方案在第一状态 θ_1 时只能获得 6.13 百万元，但乐观的决策者认为自己有能力在一定程度上改变销售状态，争取好的前景。

上述例子实质上是一种收益矩阵的决策，但是，如果是损失矩阵，就应采用"最小最小"决策准则。

【例 11 - 4】 例如，某仓储公司打算改建某集装箱货运站。它有 4 个行动方案 d_1、d_2、d_3 和 d_4 可选择，并有 4 个自然状态 θ_1、θ_2、θ_3 和 θ_4 与其相对应。但这 4 个自然状态的概率决策者无法知道，它们相应的改建费用如表 11 - 8 所示。试问选择哪一个方案为最佳？

表 11 - 8 某集装箱货运站改建费用 单位：百万元

费用 状态 方案	θ_1	θ_2	θ_3	θ_4
d_1	11	8	8	5
d_2	9	10	7	11
d_3	6	12	10	9
d_4	7	6	12	10

对于这个案例，决策者可先编制损失矩阵表，如表 11 - 9 所示。方案在各种自然状态下的最小费用为

$$\min\{11, 8, 8, 5\} = 5$$
$$\min\{9, 10, 7, 11\} = 7$$
$$\min\{6, 12, 10, 9\} = 6$$
$$\min\{7, 6, 12, 10\} = 6$$

表 11-9 损失矩阵表 单位：百万元

费用 状态 方案	θ_1	θ_2	θ_3	θ_4	$\min\limits_{\theta_j}[L_{ij}]$
d_1	11	8	8	5	5
d_2	9	10	7	11	7
d_3	6	12	10	9	6
d_4	7	6	12	10	6
决 策	$\min\limits_{d_i}\{\min\limits_{\theta_j}[L_{ij}]\}$				5

然后，求出各最小费用值中的最小值：

$$\min_{d_i}\{\min_{\theta_j}[L_{ij}]\} = \min\{5, 7, 6, 6\} = 5(\text{百万元})$$

最小值 5 百万元所对应的行动方案是 d_1，故决策者选择 d_1 方案为最佳决策方案。

11.3 赫威斯决策准则

11.3.1 赫威斯决策含义

赫威斯决策准则又称为乐观系数决策准则，它是介于悲观决策与乐观决策之间的一种决策准则，其特点是对客观条件的估计既不那么乐观，但也不悲观，主张折中平衡，是一种折中决策方法。在运用这种准则进行决策时，首先要确定一个乐观系数(对乐观程度的一个基本估计)。若认定情况完全乐观，则 $\alpha = 1$；若认定情况完全悲观，则 $\alpha = 0$；一般情况下，则 $0 < \alpha < 1$，也就是说，α 是介于 0 和 1 之间的某一个数值。

采用这种决策方法，其决策公式如下：

设一不确定型决策问题，备选方案为 $d_i(i = 1.2, \cdots, m)$，自然状态有 n 种(出现概率未知)，损益值为 $L_{ij}(i = 1, 2, \cdots, m;\ j = 1, 2, \cdots, n)$。若令：

$$f(d_i)=\alpha(\max_{\theta_j}[L_{ij}])+(1-\alpha)(\min_{\theta_j}[L_{ij}])$$

其中，$0\leqslant\alpha\leqslant 1$，则满足：

$$f(d_*)=\max_{d_i} f(d_i)$$

的方案 d_* 为 α 系数决策的最优方案。

如果所讨论的决策问题属于损失矩阵，则

$$f(d_i)=\alpha(\min_{\theta_j}[L_{ij}])+(1-\alpha)(\max_{\theta_j}[L_{ij}])$$
$$f(d_*)=\min_{d_i} f(d_i)$$

$f(d_i)$是较“最大最大”准则和“最大最小”准则更为接近实际可能情况的 d_i 方案的损益值，可称为现实估计值。

11.3.2 赫威斯决策的应用

下面举两个例子，分别说明赫威斯决策方法在收益矩阵和损失矩阵情况下的应用过程。

【例 11－5】 某港口机械厂预备生产一种新型港口机械，根据市场需求分析和估计，产品销路可分为三种状态：θ_1，销路好；θ_2，销路一般；θ_3，销路差。可供选择的行动方案也有三种：d_1，大批量生产；d_2，中批量生产；d_3，小批量生产。根据产量多少和销售情况，工厂的盈利情况也有所不同，可能获利也可能亏损，将此数值称为损益值。获利时称为收益值；亏损时称为损失值，用负号表示。本例的损益值如表 11－10 所示。试用赫威斯决策方法做出决策。

表 11－10 新型港口机械的损益值 单位：百万元

损益值 自然状态 / 方案	销路好 Q_1	销路一般 Q_2	销路差 Q_3
大批量生产 d_1	30	23	−15
中批量生产 d_2	25	20	0
小批量生产 d_3	12	12	12

在该例中，根据实际情况，决定取 $\alpha=0.6$，则 $1-\alpha=0.4$，从而可计算得出各方案的现实估计收益值如下：

$$\begin{aligned}f(d_1)&=0.6\times[\max(30,\ 23,\ -15)]+0.4[\min(30,\ 23,\ -15)]\\&=0.6\times 30+0.4\times(-15)\end{aligned}$$

$$
\begin{aligned}
&= 18 - 6 \\
&= 12(\text{万元})
\end{aligned}
$$

$$
\begin{aligned}
f(d_2) &= 0.6 \times [\max(25, 20, 0)] + 0.4 \times [\min(25, 20, 0)] \\
&= 0.6 \times 25 + 0.4 \times 0 \\
&= 15 + 0 \\
&= 15(\text{万元})
\end{aligned}
$$

$$
\begin{aligned}
f(d_3) &= 0.6 \times [\max(12, 12, 12)] + 0.4 \times [\min(12, 12, 12)] \\
&= 0.6 \times 12 + 0.4 \times 12 \\
&= 7.2 + 4.8 \\
&= 12(\text{万元})
\end{aligned}
$$

这些收益值中的最大者为

$$f(d_*) = \max(12, 15, 12) = 15 = f(d_2)$$

所以,最优方案为 d_2,即采用中批量生产方案为最佳方案。

以上讨论了收益矩阵情况下 α 系数决策方法的应用过程。如果我们根据上例,从相反的方面来提出问题,就可以构成一个损失矩阵。

例如,在上例中,若事先决策方案选 d_1,则出现销路好状态时获利最大,不会出现机会损失;若事先决策方案选 d_2,则出现销路好状态时,由于未采取方案 d_1 而带来的机会损失为 $30-25=5$(万元);若事先决策方案选 d_3,则出现销路好的机会损失为 $30-12=18$(万元)。同理,出现销路一般状态时,事先决策方案选 d_1 没有机会损失,事先决策方案选 d_2 的机会损失为 $23-20=3$(万元),事先决策方案选 d_3 的机会损失为 $23-12=11$(万元)。出现销路差状态时,事先决策方案选 d_1,的机会损失为 $12-(-15)=27$(万元),事先决策方案选 d_2 的机会损失为$12-2=12$(万元),事先决策方案选 d_3 的机会损失为 0。机会损失矩阵如表 11-11 所示。现要求在这三个方案中选择一个最佳方案。

表 11-11 机会损失矩阵 单位:百万元

机会损失 \ 自然状态 方案	销路好 Q_1	销路一般 Q_2	销路差 Q_3
大批量生产 d_1	0	0	27
中批量生产 d_2	5	3	12
小批量生产 d_3	18	11	0

现确定乐观系数 $\alpha=0.6$,则 $1-\alpha=1-0.6=0.4$。根据损失矩阵情况下赫威

斯决策公式，有：

$$\begin{aligned} f(d_1) &= 0.6\times[\min(0,\ 0,\ 27)]+0.4\times[\max(0,\ 0,\ 27)] \\ &= 0.6\times 0+0.4\times 27 \\ &= 10.8(\text{万元}) \end{aligned}$$

$$\begin{aligned} f(d_1) &= 0.6\times[\min(5,\ 3,\ 12)]+0.4\times[\max(5,\ 3,\ 12)] \\ &= 0.6\times 3+0.4\times 12 \\ &= 6.6(\text{万元}) \end{aligned}$$

$$\begin{aligned} f(d_3) &= 0.6\times[\min(18,\ 11,\ 0)]+0.4\times[\max(18,\ 11,\ 0)] \\ &= 0.6\times 0+0.4\times 18 \\ &= 7.2(\text{万元}) \end{aligned}$$

这些损失值中的最小值为

$$f(d_*) = \min(10.8,\ 6.6,\ 7.2) = 6.6 = f(d_2)$$

所以，最优方案为 d_2，即采用中批量生产的机会损失最小。

11.4　最小最大后悔值决策准则

11.4.1　最小最大后悔值决策的基本原理

在不确定型决策问题中，虽然各种自然状态的出现概率无法估计，但决策一经做出并付诸实施，必然处于实际出现的某种自然状态之中。若所选方案不如其他方案好，决策者就会感到后悔。所谓后悔值，就是所选方案的收益值与该状态下真正的最优方案的收益值之差。显然，后悔值越小，所选方案就越接近最优方案。

因此，在不确定型决策中，决策者可以在决策前计算出方案在不同自然状态下的后悔值，即先求出每种自然状态下的最大收益值与该自然状态下的其他收益之差，然后分别找出各方案对应不同自然状态下的后悔值中的最大值，最后从这些最大后悔值中找出最小的最大后悔值，将其对应的方案作为最优方案。

设某一不确定型决策，其备选方案为 $d_1,\ d_2,\ \cdots,\ d_m$，自然状态为 $\theta_1,\ \theta_2,\ \cdots,\ \theta_n$，损益值为 $L_{ij}\,(i=1,\ 2,\ \cdots,\ m;\ j=1,\ 2,\ \cdots,\ n)$。在 θ_j 状态下，必有一个方案的收益值最大，这个最大收益值可表示为

$$\max L_{ij} = \max(L_{1j},\ L_{2j},\ \cdots,\ L_{mj})$$

则在这一状态下各方案的后悔值为

$$d_1:\ \max_i L_{ij} - L_{1j}$$

$$d_2: \max_j L_{ij} - L_{2j}$$

$$\cdots\cdots$$

$$d_m: \max_i L_{ij} - L_{mj}$$

同样道理，在另一种自然状态下，各备选方案又都分别有一个后悔值。n 种自然状态，对应有 n 种后悔值。某一方案 d_i 的 n 种后悔值中的最大者，称为该方案的最大后悔值。若用 $G(d_i)$ 表示 (d_i) 方案中的最大后悔值，则

$$G(d_i) = \max_j(\max_{i=1,2,\cdots,m} L_{ij} - L_{ij})$$

对于每一个方案来说，都各有一个这样的最大后悔值，故 m 个方案就共有 m 个最大后悔值，m 个最大后悔值中的最小者，即

$$\min_{i=1,2,\cdots,m} G(d_i)$$

其对应的方案，就是"最小的最大后悔值"决策的最优方案。

11.4.2 最小最大后悔值决策的应用

【例 11-6】 某码头决定扩展港口装卸业务，有下列三个方案可供选择：d_1，建立新泊位大幅度提高装卸能力；d_2，改造原有泊位使装卸能力达到中等水平；d_3，利用原有泊位先做试点。船公司对该码头扩展后的需求情况有以下 4 种可能状态：θ_1，需求量很大，停靠船舶大量增加；θ_2，需求量较大，停靠船舶适量增加；θ_3，需求量不大，停靠船舶数量增加幅度很小；θ_4，需求量很小，停靠船舶数量没有增加。3 种方案和 4 种状态下的损益值如表 11-12 所示。试求其最佳方案。

表 11-12 某产品的损益值

损益值 / 自然状态 / 方案	θ_1	θ_2	θ_3	θ_4
建新泊位 d_1	80	40	−30	−70
改造原有泊位 d_2	55	37	−15	−40
利用原有泊位 d_3	31	31	9	−1

这是一个不确定型决策问题，备有 3 个方案、4 种自然状态。由于：

$$\max_{i=1,2,3} L_{i1} = \max(80, 55, 31) = 80$$

$$\max_{i=1,2,3} L_{i2} = \max(40, 37, 31) = 40$$

$$\max_{i=1,2,3} L_{i3} = \max(-30, -15, 9) = 9$$

$$\max_{i=1,2,3} L_{i4} = \max(-70, -40, -1) = -1$$

所以,方案的最大后悔值为

$$\begin{aligned} G(d_1) &= \max[80-80, 40-40, 9-(-30), (-1)-(-70)] \\ &= \max(0, 0, 39, 69) \\ &= 69 \end{aligned}$$

$$\begin{aligned} G(d_2) &= \max[80-55, 40-37, 9-(-15), (-1)-(-40)] \\ &= \max(25, 3, 24, 39) \\ &= 39 \end{aligned}$$

$$\begin{aligned} G(d_3) &= \max[80-31, 40-31, 9-9, (-1)-(-1)] \\ &= \max(49, 9, 0, 0) \\ &= 49 \end{aligned}$$

最优方案按下式决定:

$$\min_{i=1,2,3} G(d_i) = \min(69, 39, 49) = 39 = G(d_2)$$

计算结果表明,方案 d_2 即改建现有泊位达到中等装卸水平为最佳方案。

11.5 等概率决策准则

等概率决策准则是由拉普拉斯(Laplace)提出的一种决策准则。运用这种准则进行决策时,应做这样一个假定,即当决策者在决策过程中不能肯定哪种状态容易出现、哪种状态不容易出现时,就"一视同仁",认为它们出现的可能性(概率)是相等的。在这种假定条件下再利用同等概率来计算各个行动方案的期望收益值,具有最大收益期望值的方案就是最优方案。同样,如果我们计算的是同等概率条件下的损失值,那么具有最小损失值的方案就是最优方案。

【例 11-7】 为了适应市场需要,某港口集团公司提出了扩大再生产的 3 种方案:① 对原码头进行扩建;② 对原码头进行设备技术改造;③ 建设新码头。每年的利润(百万元)和市场需求情况如表 11-13 所示。

表 11-13 某码头生产利润或亏损表 单位:百万元

扩大生产方案	自然状态		
	市场需求大 Q_1	市场需求一般 Q_2	市场需求差 Q_3
扩建码头 d_1	15	13	−4
原码头技术改造 d_2	8	7	4
建新码头 d_3	17	12	−6

假定三种自然状态(市场需求大、一般、市场需求差)发生的概率相等,均为1/3,试用等概率决策准则进行决策。

先计算出各种方案的等概率期望收益值,如表 11 - 14 所示。

表 11 - 14　等概率期望收益值计算表　　单位:百万元

行动方案	等概率期望收益值 $E(d_i)$
扩建码头 d_1	$\frac{1}{3}\times 15+\frac{1}{3}\times 13+\frac{1}{3}\times(-4)=8$
原码头技术改造 d_2	$\frac{1}{3}\times 8+\frac{1}{3}\times 7+\frac{1}{3}\times 4=6.33$
建新码头 d_3	$\frac{1}{3}\times 17+\frac{1}{3}\times 12+\frac{1}{3}\times(-6)=7.67$

从表 11 - 14 计算结果可以看出,扩建码头方案的期望收益值 8 百万元为最大。因此,按等概率准则进行决策,扩建码头方案为最优方案。

11.6　决策准则的评价与选择

11.6.1　各种决策方法的比较

决策者利用表 11 - 15 的资料,采用在不确定型条件下的各种决策方法,可以得到不同的最优方案,如表 11 - 16 所示。出现这种情况的原因是由于每一种决策方法都是考虑了决策者的决策心理、感情和愿望而制定的。

表 11 - 15　各个方案的企业年损益值　　单位:万元

方案 \ 利润 \ 自然状态	需求量较高	需求量一般	需求量较低
第一方案	70	30	50
第二方案	20	20	90
第三方案	50	80	40
第四方案	40	100	20

表 11-16 各种不同决策方法结果比较

采用不同的决策方法	选用的最佳方案
1. "好中求好"决策方法	第四方案
2. "坏中求好"决策方法	第一方案
3. 赫威斯决策方法（$\alpha=0.7$）	第四方案
4. "最小的最大后悔值"决策方法	第二方案

对于解决不确定型决策问题，现在在理论上还不能证明哪一种评选标准是最合理的。因此，在实际工作中究竟采用哪一种决策方法，还带有相当程度的主观随意性。国外有许多学者还在继续讨论这个问题。一般来说，如果要把各种决策方法做出比较，那么，"最大最小"决策方法主要由那些比较保守稳妥并害怕承担较大风险的决策者所采用，"最大最大"决策方法主要由那些对有利情况的估计比较有信心的决策者所采用，赫威斯决策方法主要由那些对形势判断既不乐观也不太悲观的决策者所采用，"最小最大后悔值"决策方法主要由那些对决策失误的后果看得较重的决策者所采用。

11.6.2 各种决策方法在应用时的选择

表 11-16 的计算结果已表明，不同的决策方法会导致不同的最优方案。很明显，这里有正确的决策结果，也有错误的决策结果。这就提出了一个问题：根据什么选取决策方法，决策结果才是正确可靠的？决策方法的选择固然与决策者的主观意志有关。但是，决策方法的选择又不完全取决于决策者的主观意志。如果一个决策者不顾决策问题所面临的客观环境，仅凭想当然和个人兴趣选取决策方法，那该决策者所做出决策的可靠性是值得怀疑的。可以想象，如果同一个问题处于同一个客观环境中，可以采用不同决策方法进行分析，那么各种决策方法都将失去其存在的意义。事实上，各种决策方法都具有一定的假定条件，而任何一种假定条件都无法概括现实过程中错综复杂的关系。因此，决策方法的选择只能以决策问题所处的客观条件为基础。下面我们用实例说明决策方法的选择问题。

例如，某集装箱装卸公司根据往年船公司对该码头集装箱船舶装卸的需求量预测，制订了以下 3 个码头生产调度系统开发目标方案：

(1) 全面引进技术，进口码头生产调度系统。

(2) 全部依靠自己的力量，改造现有系统，实现决策目标。

(3) 自行改造为主，技术引进为辅。

该码头首先对 3 个方案进行了定性分析，并认为：① 采用第一方案的优点是技术先进，可以提高码头生产调度信息传递的速度进而提高生产效率；缺点是耗资

大。② 采用第二方案的优点是费用少；缺点是周期长，受技术条件限制，开发后的系统不易达到国际先进水平。③ 采用第三方案的优点是关键技术和设备可达到世界先进水平，周期短，投资不多，而且可以承担自行改造为主的任务；缺点是生产能力没有第一方案大。

定量分析是定性分析的深化，是决策过程中不可缺少的环节。进行了定性分析后，还要进行定量分析。

根据该码头的有关资料，得到了如表 11-17 所示的损益矩阵表。

表 11-17　某股份公司损益矩阵表　　单位：万元

利润 / 自然状态 / 方案	高需求 Q_1	中需求 Q_2	低需求 Q_3
全面引进 d_1	44 040	37 592	31 300
全部自制 d_2	36 450	35 450	34 500
改进与改造相结合 d_3	43 840	40 592	34 300

方案的选择过程如下：

(1) 按照“最大最大”决策方法，最高利润为 44 040 万元，它所对应的方案为 d_1，即全面引进，对应的自然状态是高需求。但是，根据当前航运市场和全球经济发展形势，最近 1 年内出现最高船舶装卸需求峰值的可能性不大。另外，全面引进需要 400 万美元外汇，该码头不具备这种条件。因此，该问题不能按照“最大最大”决策方法进行决策。

(2) 按照“最大最小”决策方法，34 500 万元是决策目标值，它所对应的方案为 d_2，即全部自制，对应的自然状态是低需求。但是，中国总体经济发展形势是乐观的，低需求出现的可能性很小。同时，采用全部自制方案将造成生产能力无潜力，码头生产效率很难达到国际先进水平。因此，也不可按照“最大最小”决策方法进行决策。

(3) 根据预测资料以及该码头的客观条件，该公司认为：决策因素中既有乐观的一面，也有悲观的一面，且乐观因素大于悲观因素。因此，按照赫威斯决策方法进行决策。经分析，取乐观系数为 $\alpha = 0.7$，则 $1-\alpha = 1-0.7 = 0.3$。其计算过程和结果如下：

$$
\begin{aligned}
f(d_1) &= 0.7 \times [\max(44\,040,\ 37\,592,\ 31\,300)] + \\
&\quad\ 0.3 \times [\min(44\,040,\ 37\,592,\ 31\,300)] \\
&= 0.7 \times 44\,040 + 0.3 \times 31\,300 = 30\,828 + 9\,390 \\
&= 40\,218(\text{万元})
\end{aligned}
$$

$$\begin{aligned}f(d_2) &= 0.7\times[\max(36\,450, 35\,450, 34\,500)]+\\&\quad 0.3\times\min(36\,450, 35\,450, 34\,500)\\&= 0.7\times 36\,450+0.3\times 34\,500\\&= 25\,515+10\,350\\&= 35\,865(\text{万元})\\f(d_3) &= 0.7\times[\max(43\,840, 40\,592, 34\,300)]+\\&\quad 0.3\times[\min(43\,840, 40\,592, 34\,300)]\\&= 0.7\times 43\,840+0.3\times 34\,300\\&= 30\,688+10\,290\\&= 40\,978(\text{万元})\end{aligned}$$

这些利润值中的最大者为

$$f(d_*) = \max(40\,218, 35\,865, 40\,978) = 40\,978 = f(d_3)$$

计算结果表明，方案 d_3 即引进与改造相结合方案为最佳方案。

（4）由于赫威斯决策方法中 α 的取值带有一定的主观性，因此，决定再用“最小最大后悔值”决策方法对所选方案进行验证。

其计算过程和计算结果如下：

$$\begin{aligned}\max_{i=1,2,3} L_{i1} &= \max(44\,040, 36\,450, 43\,840) = 44\,040\\\max_{i=1,2,3} L_{i2} &= \max(37\,592, 35\,450, 40\,592) = 40\,592\\\max_{i=1,2,3} L_{i3} &= \max(31\,300, 34\,500, 34\,300) = 34\,500\end{aligned}$$

所以，各方案的最大后悔值为

$$\begin{aligned}G(d_1) &= \max(44\,040-44\,040, 40\,592-37\,592, 34\,500-31\,300)\\&= \max(0, 3\,000, 3\,200) = 3\,200\\G(d_2) &= \max(44\,040-36\,450, 40\,592-35\,450, 34\,500-34\,500)\\&= \max(7\,590, 5\,142, 0) = 7\,590\\G(d_3) &= \max(44\,040-43\,840, 40\,592-40\,592, 34\,500-34\,300)\\&= \max(200, 0, 200) = 200\end{aligned}$$

最优方案按下式决定：

$$\min_{i=1,2,3} G(d_i) = \min(3\,200, 7\,590, 200) = 200 = G(d_3)$$

计算结果表明，按“最小最大后悔值”决策方法，其最小值 200 万元所对应的是方案 d_3，因此，引进与改造相结合的方案为最佳方案。这个结论同用赫威斯决策

方法所得出的结论一致。

最后，该股份公司决定采用引进与改造相结合的方案。

思考与练习

1. 名词解释：最大最小决策准则，最大最大决策准则，赫威斯决策准则。

2. 港口装卸码头，在市场装卸需求不稳定、竞争激烈、市场预测未来前景不够理想的情况下，决定按最大最小准则决策采用较稳的运营方针。该公司的经营状况、市场状态即年收益情况如题表 11－1 所示：

题表 11－1 单位：百万元

收益值 市场状态 / 经营方式	装卸需求大	一般	装卸需求小
α_1	9	7	4
α_2	12	8	−1
α_3	10	6	3

问公司决策者按此准则应采用何种经营方式？

3. 某企业为了适应市场日益更新的需要，决定投产一种新产品。为此，提出了 3 种备选方案：D_1，引进国外生产线；D_2，对原生产线进行技术改造；D_3，与国内某同类企业进行合作生产。但该企业对该种新产品的投产又感到不能盲目乐观，决定以 $\alpha=0.65$ 的乐观系数，按照赫威斯准则决策。该企业估算以上 3 种方案在市场出现高需求、中等需求、低需求的情况下 10 年之内所获利润(单位：万元)如题表 11－2 所示。问该企业应该选择哪一种方案为最优方案？

题表 11－2

市场情况 / 方案	高需求	中等需求	低需求
D_1	650	200	−150
D_2	400	180	100
D_3	350	270	5

4. 某码头建设问题的损益值如题表 11－3 所示：

题表 11-3 单位：万元

决策方案	自然状态	
	装卸量大	装卸量小
建设大型泊位 D_1	200	−20
建设中型泊位 D_2	150	20
建设小型泊位 D_3	100	60

要求：

(1) 按最大最大决策准则选择一个决策方案。

(2) 按最大最小决策准则选择一个决策方案。

(3) 按赫威斯决策准则选择一个决策方案。

(4) 按最小最大后悔值法选择一个决策方案。

12 风险型决策方法

12.1 风险型决策概述

12.1.1 风险型决策的定义

风险型决策是决策者根据几种不同自然状态可能发生的概率所进行的决策。决策者所采取的任何一个行动方案都会遇到一个以上自然状态所引起的不同结果,这些结果出现的机会是用各种自然状态出现的概率来表示的。由于不管决策者选择哪个行动方案,都要承担一定风险,因而又称为随机型决策。

【例 12-1】 某建筑公司承建一项码头工程,需要决定下个月是否开工。如果开工后天气好,可以按期完工,就可获得利润 5 万元;如果开工后天气坏,将造成损失 2 万元;如果不开工,不管天是好还是坏,都要付出误工损失 5 千元。根据历史气象统计资料,预测下月天气好的概率是 0.4,天气坏的概率是 0.6,追求目标利润最大损失最小,该公司应决定开工还是不开工?

本例内容列表如 12-1 所示。

表 12-1 工程开工决策表

自然状态 / 状态概率 / 行动方案	天气好	天气坏
状态概率	0.4	0.6
开　工	50 000	−20 000
不开工	−5 000	−5 000

这是一个风险型决策问题。因为不论选择哪个方案,都要承担一定的风险。选择开工方案可能遇上天气坏,选择不开工方案又可能遇上天气好,都会蒙受损失。决策的任务就是要争取获得最大的可能利润。

12.1.2 风险型决策需具备的条件

(1) 存在着决策者希望达到的一个或一个以上明确的决策目标。最常用的决策目标是要求获得最大的利润。

(2) 存在着决策者可以主动选择的两个以上的行动方案，即存在两个以上决策变量。

(3) 存在着不以决策者的主观意志为转移的两种以上的自然状态，即存在着两种以上的状态变量。

(4) 存在着可以具体计算出来的不同行动方案在不同自然状态下的损益值。

(5) 存在着决策者可以根据有关资料事先估计或计算出来的各种自然状态将会出现的概率。

12.1.3 风险型决策所使用的概率

由于使用了概率，所以风险型决策也属于不确定型的决策，运用什么样的概率和概率值的准确程度，是做好风险型决策的至关重要的问题。风险型决策所使用的概率有以下几种：

1) 客观概率

客观概率是根据事件过去和现在的资料所确定或计算的某个事件出现的概率。在客观概率中又有先验概率与后验概率之分。前者是根据事件的历史资料来确定，后者是根据历史资料和现实资料计算而获得的。利用后验概率决策要比运用先验概率决策准确可靠一些。

2) 主观概率

主观概率是由决策者主观判断所确定的某个事件出现的概率。这种概率没有事件过去或现在的资料作为实证依据。决策者一般是根据以往的表现和经验结合当前形势动态来大致确定的。当然，这与决策者个人的智慧、经验、胆识、个性等有密切关系。在一般情况下，主观概率不及客观概率准确可靠。

12.2 期望损益决策法

期望损益决策法是以损益期望值为基础，将不同方案的期望值相互比较，选择期望收益值最大或期望损失值最小的方案为最优方案。这种方法包括期望收益决策法和期望损失决策法。

12.2.1 期望收益决策法

期望收益决策是以不同方案的期望收益作为择优的标准，选择期望收益最大的方案为最优方案。

【例 12-2】 如【例 12-1】中，根据表 12-1 的情况，是决定开工还是不开工？

【解】 (1) 计算各方案的期望收益值。

开工方案：

$$0.4\times 50\,000+0.6\times(-20\,000)=8\,000(\text{元})$$

不开工方案：

$$(0.4+0.6)\times(-5\,000)=-5\,000(\text{元})$$

(2) 决策。

根据计算结果，开工方案能够获利 8 000 元，如选择不开工方案则损失 5 000 元，因此，选择开工方案是决策最优方案。

【例 12-3】 某港口机械零配件销售商要拟订 6 月、7 月、8 月某港口机械零配件的日进货计划。配件进货成本为每个 60 元，销售价格为 110 元，即当天能卖出去，每个可获利 50 元，如果当天卖不出去，剩余一个就要由于仓储保管及其他原因而亏损 20 元。现市场需求情况不清楚，但有前两年同期计 180 天的日销售资料，如表 12-2 所示。问应怎样拟订零配件的日进货计划，才能使利润最大？

【解】 (1) 根据前两年同期日销售量资料进行统计分析，确定不同日销售量的概率，如表 12-2 所示。

表 12-2 港机零配件日销售概率表

日销售量(个)	完成日销售量的天数	概　率
50	36	36/180=0.2
60	72	72/180=0.4
70	54	54/180=0.3
80	18	18/180=0.1
$\sum$	180	1

(1) 计算每天可能的日收益，如表 12-3 所示。

表 12-3 港机零配件不同进货方案的收益表

日进货量/个 \ 条件利润/元 \ 状态概率 \ 日销售量/个	50	60	70	80	期望利润
	0.2	0.4	0.3	0.1	
50	2 500	2 500	2 500	2 500	2 500
60	2 300	3 000	3 000	3 000	2 860
70	2 100	2 800	3 500	3 500	2 940
80	1 900	2 600	3 300	4 000	2 810

各个进货方案在不同的日销售量条件下的利润是随供需关系而定的。设以 Q 代表日进货量，以 D 代表市场的日可能销售量，则每日条件利润的计算方法如下：

$$当\ Q \leqslant D\ 时，条件利润 = (110-60)Q = 50Q$$

$$当\ Q > D\ 时，条件利润 = (110-60)Q - 20(Q-D) = 70D - 20Q$$

例如，日进货 50 个，售出 50 个，即 $Q=D$，条件利润为 $50\times50=2\,500$ 元，若需求量大于 50 个，利润仍为 2 500 元；又日进货 60 个而售出 50 个，即 $Q>D$，这时条件利润为 $70\times50-20\times60=2\,300$ 元，如表 12-3 所示。

(3) 计算各个进货方案的期望利润值。各个方案的期望利润，是在收益表的基础上，将每个方案在不同自然状态下的利润值乘以该自然状态发生的概率值之和。

例如，进货 60 个方案的期望利润为

$$2\,300\times0.2+3\,000\times0.4+3\,000\times0.3+3\,000\times0.1=2\,860(元)$$

4 个方案的期望利润值列于表 12-3 中。

(4) 决策。从表 12-3 的计算结果可以看出：进货 70 个的计划方案的期望利润为最大。因此，该店的最优进货方案是日进货 70 个零配件。

应该指出，上述的期望利润是不同进货方案在各种不同市场需求状态下，用销售概率对条件利润加权计算，可能获得的一种平均利润。它既具有一定的代表性，也具有一定的风险性。如能够加强市场销售趋势的调查研究，掌握完整的市场情况资料，使每天的进货量完全符合市场的需要量，既无缺货，又无剩余，这实际上已属于确定型决策，这时可以获得最大的利润。情况如表 12-4 所示。

表 12-4 具有完整资料情况下的最大利润表

日销售量/个 状态概率 日进货量/个　条件利润/元	50	60	70	80
	0.2	0.4	0.3	0.1
50	2 500	——	——	——
60	——	3 000	——	——
70	——	——	3 500	——
80	——	——	——	4 000

根据表 12-4 的资料，可算得具有完整资料情况下的期望利润为

$$2\,500\times0.2+3\,000\times0.4+3\,500\times0.3+4\,000\times0.1=3\,150$$

这期望利润值比表 12-3 中没有完整资料情况下的最大期望利润 2 940 元还

要大 210 元(即 3 150－2 940＝210)。这个 210 元就是完整资料的实际价值。在决策时是否需要花费力量进行调查研究取得完整的市场资料，这就需要比较完整资料本身实际价值的大小和搜集完整资料开支费用的大小来确定。一般来说，搜集完整资料的调查费小于资料本身的价值，搜集完整资料的工作才值得进行。

12.2.2 期望损失决策法

期望损失决策法以不同方案的期望损失作为择优的标准，选择期望损失最小的方案为最优方案。

下面我们以上例的资料来说明其决策过程。

(1) 根据每天可能的销售量，编制不同进货方案的条件损失表(见表 12－5)。

表 12－5 港机零配件不同进货方案的损失表

日进货量/个 \ 条件利润/元 \ 日销售量/个 状态概率	50	60	70	80	期望损失
	0.2	0.4	0.3	0.1	
50	0	500	1 000	1 500	650
60	200	0	500	1 000	290
70	400	200	0	500	210
80	600	400	200	0	340

这些条件损失值既包括进货超过可能销售量的经营损失，也包括由于进货不足而造成的机会损失。其计算方法是：用表 12－3 中每列的最大收益值减去该列中其他各种收益值。

表内条件损失值以“0”为对角线，左上部为机会损失，右下部为经营损失。

(2) 计算各个进货方案的期望损失值。期望损失的计算方法与期望收益相同，是以各方案在不同自然状态下的损失值乘以其概率值之和。其计算结果列于表 12－5 中。

(3) 决策。从表 12－5 的计算结果可以看出：进货 70 个的计划方案的期望损失为最小，因此以最小期望损失为标准，最优进货方案仍是进货 70 个零配件。这个决策结论与以最大期望收益为标准所做出的结论相同。

最后指出：根据表 12－4 算出的具有完整资料情况下的最大期望利润 3 150 元，正好等于表 12－3 中的最大期望利润 2 940 元加上表 12－5 中最小期望损失 210 元。这样便把表 12－3、表 12－4 和表 12－5 中的计算结果有机地联系起来了。现将这种联系列于表 12－6 中。

表 12-6 期望损益决策指标联系表

日进货量/箱	期望利润/元	期望损失/元	完整资料情况下的期望利润/元
(1)	(2)	(3)	(4)=(2)+(3)
50	2 500	650	3 150
60	2 860	290	3 150
70	2 940	210	3 150
80	2 810	340	3 150

12.3 边际分析决策法

上节介绍的期望损失决策法，在行动方案和自然状态较小时，计算使用都很方便，是一种较好的经济决策法。但是，当行动方案和自然状态较多时，仍使用期望损益决策法就会使计算工作过于繁重。因此，若决策问题满足：① 行动方案(决策变量)和自然状态(状态变量)均可以用有序的数量表示；② 决策后果的损益值是决策变量与状态变量的线性函数。这时可采用边际分析法进行决策，这种方法比较简便，能加速决策过程。港口机械零配件进销的例子能满足上述两个条件，因此也可以采用边际分析法来做出决策。

边际分析决策法又称增量分析决策法。下面就以港机零配件进货为例说明这种分析方法。

当我们分析该销售商进货应安排多少个为最佳时，从边际分析入手，就要考虑到，每增加进货 1 个，必然要出现两种可能：当天顺利卖出去或卖不出去。当天顺利卖出去可以多得利润 50 元，这个 50 元称为边际利润(marginal profit，MP)；未能卖出去将蒙受损失 20 元，这个 20 元称为边际损失(marginal loss，ML)。由于进货每增加 1 个后能否卖出去取决于市场的需要量，而在风险型的情况下，市场的可能需要只能以销售概率来表示。因此，是否增进 1 个，这不仅要考虑增进 1 个的边际利润和边际损失各是多少，而且要考虑到销售概率。即要比较它的期望边际利润和期望边际损失的大小。若期望边际利润大，说明有利可图，则应进货；若期望边际损失大，说明将蒙受损失，则不应进货；当期望边际利润等于期望边际损失时，说明已达到平衡，这就是最大进货界限，若再增加进货，就会出现损失大于利润的现象。以上就是是否增加进货的决策标准。根据这一标准，我们可以推导出最佳的日进货量。但这里还得引进一个“累积概率”的概念，累积销售概率的意思就

是至少能销售出某一数量的概率。如本例中，至少能销售出 60 个的概率，就包括售出 60 个、70 个、80 个的概率之和，即等于 0.4+0.3+0.1=0.8，同理，可计算出各销售量的累积概率，如表 12-7 所示。

表 12-7 港机零配件日销售量的积累销售概率表

日销售/个	销售概率	累计销售概率
50	0.2	1
60	0.4	0.8
70	0.3	0.4
80	0.1	0.1

设以 P 代表新增进一个零配件能顺利售出的累积概率，新增进一个零配件不能售出的概率为 $1-P$。根据期望边际利润等于期望边际损失的决策准则，就应有：

$$P \times MP = (1-P) \times ML$$

化简可得

$$P = \frac{ML}{MP+ML} \tag{12-1}$$

这个 P 就是保证新增进一个零配件能卖出去的最低概率，称为转折概率。而累积销售概率等于转折概率 P 的日销售量，就是利润期望值最大的日进货量，也就是最佳的进货量。

归纳以上的讨论，边际分析决策的步骤是：

(1) 算出转折概率。

(2) 编制各种自然状态的累积概率表。

(3) 决策。

在累积概率表中找出与转折概率相对应的自然状态，这个自然状态就是最佳行动方案。

【例 12-4】 用边际分析法对【例 12-3】中的进货计划进行决策。

【解】 (1) 计算转折概率。已知 $MP=50$，$ML=20$，由式(12-1)可得转折概率为

$$P = \frac{20}{50+20} = \frac{2}{7} \approx 0.286$$

(2) 编制各日销售量的累积销售概率表,如表 12-7 所示。

(3) 决策。在表 12-7 中,没有累积概率正好等于 0.286,而 0.286 介于 0.4 与 0.1 之间,即最佳日进货量也应介于 70 到 80 箱之间。我们可以用线性内插法计算出具体的最佳进货量数字:

$$\text{最佳进货量} = 70 + \frac{80-70}{0.4-0.1} \times (0.4-0.286) = 70 + 3.8 = 73.8(\text{箱})$$

为了进一步说明问题,现将各种进货方案下的期望边际利润和期望边际损失计算列成表 12-8 比较如下。

表 12-8 期望边际利润与期望边际损失比较表

日进货量/个	累计销售概率	期望边际利润 $P \times MP$/元	比较关系	期望边际损失 $(1-P) \times ML$/元
50	1	$1.0 \times 50 = 50$	$>$	$0 \times 20 = 0$
60	0.8	$0.8 \times 50 = 40$	$>$	$0.2 \times 20 = 4$
70	0.4	$0.4 \times 50 = 20$	$>$	$0.6 \times 20 = 12$
73.8	0.286	$0.286 \times 50 = 14.3$	$=$	$0.714 \times 20 = 14.3$
80	0.1	$0.1 \times 50 = 5$	$<$	$0.9 \times 20 = 18$

从表 12-8 中可以看出,当日进货 50、60、70 个时,期望边际利润都大于期望边际损失,由于盈利的可能性大,这时应继续进货。但是当进货增至 80 箱时,期望边际利润小于期望边际损失,发生损失的机会已大,因此进货量不宜由 70 箱增加到 80 箱。在累积销售概率为 $2/7 \approx 0.286$ 时,相应的进货量为 73.8 箱,这时期望边际利润正好与期望边际损失相等。这是一个转折点,进货超过这一点,期望边际利润便都小于期望边际损失了。所以称这时的累积概率为转折概率,相应的进货量为最佳进货量。所以,本例的最优决策方案,应是日进货 73.8 箱。

如果从现有的 4 个日进货方案中,选定一个,那么就选定日进货 70 箱为最佳进货方案。因为只有累积概率小于 0.286 的日进货量,其期望边际利润才小于其期望边际损失。这个结果与上一节的结果相同。

12.4 决策树法

决策树法是风险决策中常用的方法。它的优点是能使决策问题形象直观,思路清晰,便于思考与集体探讨。特别在多级决策活动中,能起到层次分明、一目了然、计算简便的作用。

12.4.1 决策树的结构

决策树又称为决策图，它是以方框和圆圈为结点，由直线连接而形成的一种树枝形状的结构，如图 12－1 所示。

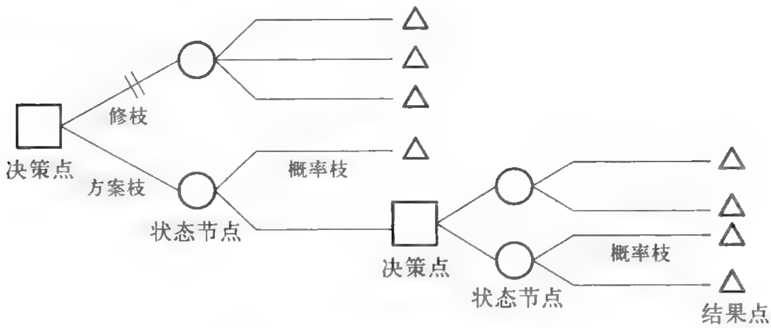


图 12－1 决策树模型

在图 12－1 中，方框结点称决策点。由决策点引出若干条直线，每条直线代表一个方案，称为方案枝。在各个方案枝的末端画上一个圆圈，称为状态结点。由于状态结点引出若干条直线，每条直线代表一个自然状态及其可能出现的概率，故称为概率枝。在概率枝末端画个三角，称为结果点。在结果点旁边列出不同状态下的收益值或损失值，以供决策之用。

一般决策问题具有多个行动方案，每个方案又常常出现多种自然状态，因此决策图形都是由左向右、由简入繁，组成一个树形的网状图。

应用决策树进行决策的过程是：由右向左逐步后退进行分析。根据右端的损益值和概率枝的概率，计算出同一方案不同自然状态下的期望收益值或损失值，然后根据不同方案的期望收益值或损失值的大小做出选择，对落选的方案在图上需要进行修枝，即在落选的方案枝上画上“//”符号，以表示舍弃不选的意思。最后决策点只留下一条树枝，即为决策中的最优方案。

12.4.2 单级决策

一个决策问题，如果只需进行一次决策就可以选出最优方案，达到决策目的，这种决策称为单级决策。下面举例加以说明。

【例 12－5】 为了适应市场的需要，某港口城市提出了扩大某个码头生产的两个方案。一个方案是建设大泊位，另一个方案是建设小泊位，两者的使用期都是10 年。建设大泊位需要投资 600 万元，建设小泊位需要投资 280 万元，两个方案的

每年损益值及自然状态的概率如表 12-9 所示。试用决策树评选出合理的决策方案。

表 12-9 年度损益值计算表 单位：万元/年

自然状态	概 率	方 案	
		建大泊位	建小泊位
装卸量大	0.7	200	80
装卸量小	0.3	−40	60

【解】 (1) 画出决策树，如图 12-2 所示。

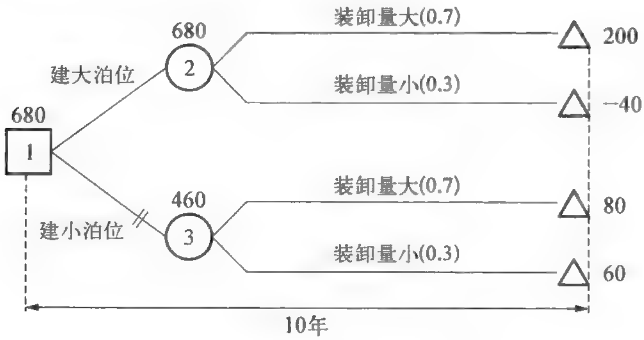


图 12-2 一 级 决 策 树

(2) 计算各点的期望损益值。

点②：$[0.7\times 200+0.3\times(-40)]\times 10-600=680$(万元)

点③：$(0.7\times 80+0.3\times 60)\times 10-280=460$(万元)

(3) 进行决策。把点②与点③的期望损益值进行比较，可知合理的决策方案是建设大泊位。

【例 12-6】 在【例 12-5】中，如果把 10 年分为前 3 年和后 7 年两期考虑。根据市场预测：前 3 年装卸量大的概率为 0.7；若前 3 年装卸量大，则后 7 年装卸量大的概率为 0.8。前 3 年装卸量小的概率为 0.3；若前 3 年装卸量小，则后 7 年装卸量小的概率为 0.9，则建大泊位和建小泊位两个方案哪个为好？

【解】 (1) 画出决策树，如图 12-3 所示。

(2) 计算各点的期望损益值。

点④：$[0.8\times 200+0.2\times(-40)]\times 7=1\,064$(万元)

点⑤：$[0.1\times 200+0.9\times(-40)]\times 7=-112$(万元)

点②：$0.7\times 200\times 3+0.7\times 1\,064+0.3\times(-40)\times 3+0.3\times(-112)-$

600 =495.2(万元)

这是建大厂的期望收益值。

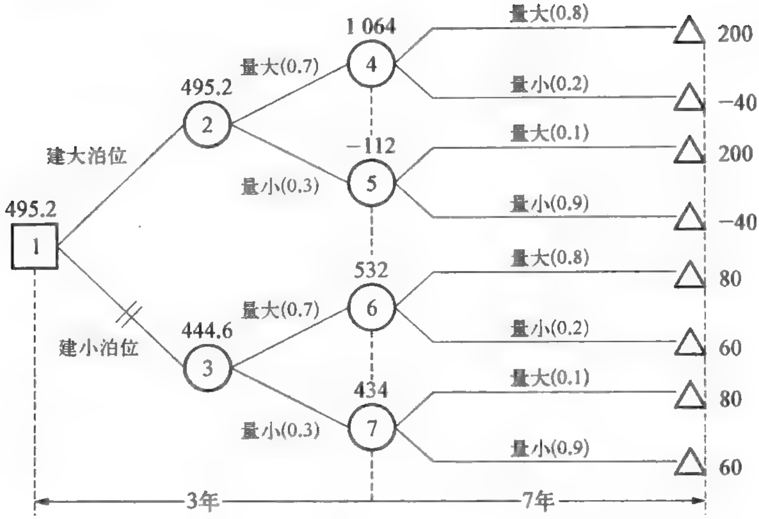


图 12-3 一级决策树

点⑥:(0.8×80+0.2×60)×7 = 532(万元)

点⑦:(0.1×80+0.9×60)×7 = 434(万元)

点③:0.7×80×3+0.7×532+0.3×60×3+0.3×434−280 = 444.6(万元)

这是建小厂的期望收益值。

(3) 进行决策。建大厂的期望收益值 495.2 万元,大于建小厂的 444.6 万元。因此,建大厂仍是最优方案。

在实际工作中,还需要把不同时期的投资额和损益值等按照复利原则折算为现值,才能进行对比研究。这些问题已属于投资决策范围了。

12.4.3 多级决策

一个决策问题如果需要进行两次或两次以上的决策才能选出最优方案、达到决策目的,这种决策称为多级决策。

【例 12-7】 假定在【例 12-6】中又提出第三方案,即先建设小泊位,如果装卸量大,则 3 年后再进行扩建。扩建投资需要 400 万元,扩建后,也可使用 7 年,每年的损益值与大泊位相同。这个方案与建大泊位方案比较,哪个方案较好?

【解】 (1) 画出决策树,如图 12-4 所示。

(2) 计算各点的期望损益值。

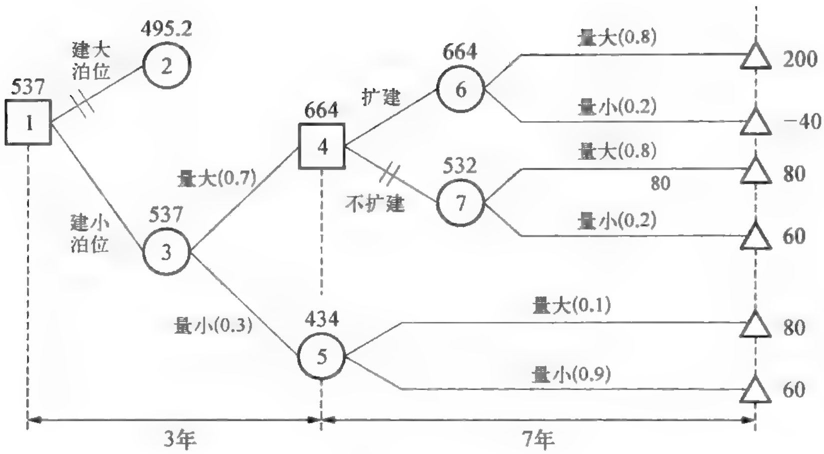


图 12-4 二级决策树

点②：495.2(万元)(计算见【例 12-5】)

点⑥：$[0.8\times200+0.2\times(-40)]\times7-400=664$(万元)

点⑦：$(0.8\times80+0.2\times60)\times7=532$(万元)

把点⑥和点⑦的期望值相比较，前者的期望收益值较大，所以应当选择扩建方案，对不扩建方案进行修枝。把点⑥的 664 万元移到点④上来，这是第一次决策。

点⑤：$(0.1\times80+0.9\times60)\times7=434$(万元)

点③：$0.7\times80\times3+0.7\times664+0.3\times60\times3+0.3\times434-280=537$(万元)

(3) 进行决策。点③的期望收益值为 537 万元，大于点②的 495.2 万元。因此，最优方案是前 3 年建小厂，如果销路好，后 7 年进行扩建的方案，而不是建大工厂的方案了。本例进行了两次决策，才选出最优方案，所以是二级决策问题。

【例 12-8】 某小型集装箱仓储公司考虑是否要投资 15 000 元开发一种仓储调度的新系统模块。如开发，可望有 60%的机会成功；如获成功，可申请专利，专利申请费为 5 000 元；获批准的机会是 50%。如试验成功，不管专利是否被批准，仓储公司都有自用还是出售这种权利的选择。在使用中，这种新系统模块可望在 4 年内有利可图；获利大小，受能否取得专利权的影响，受该工序本身的性能(能产生多少利润)的影响，还受政府的补助(用减税或津贴等方式)的影响。政府补助可分大量、中等及少量 3 种。各种情况下的条件利润如下：

(1) 出售权利的收益。批准专利权时，可得 40 000 元；未批准专利权时，可得 25 000 元；不申请专利权时，可得 30 000 元。

(2) 自己使用的每年收益。自己使用新系统模块的年收益，视政府补助大小和专利情况不同而异，如表 12-10 所示。

表 12－10　使用新系统模块利润表　　单位：元

政府补助	概　　率	批准专利	未批准专利	不申请专利
大　量	0.3	15 000	7 000	11 000
中　等	0.4	10 000	5 500	8 000
小　量	0.3	6 000	3 000	5 000

试用决策树评选出合理的决策方案。

【解】 (1) 画出决策树，如图 12－5 所示。

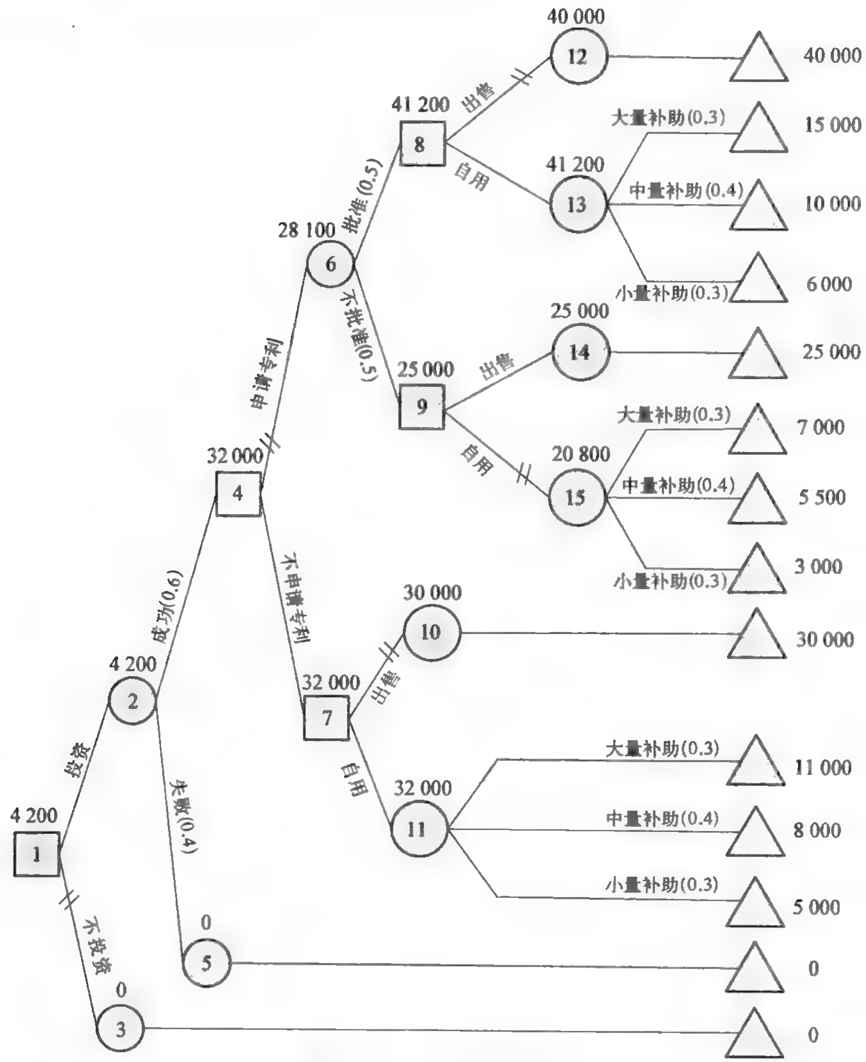


图 12－5　三 级 决 策 树

(2) 计算各点期望利润。

点⑫：40 000(元)

点⑬：$(0.3\times 15\,000+0.4\times 10\,000+0.3\times 6\,000)\times 4=41\,200$(元)

把点⑫的期望值与点⑬的相比较，后者的期望收益值为大，所以应当选择自用方案，对出售方案进行修枝。把点⑬的 41 200 元移到点⑧上来，这是申请专利权被批准后进行的第一次决策。

点⑭：25 000(元)

点⑮：$(0.3\times 7\,000+0.4\times 5\,500+0.3\times 3\,000)\times 4=20\,800$(元)

把点⑭的期望值与点⑮的相比较，前者的期望收益值较大，所以应当选择出售方案，对自用方案进行修枝。把点⑭的 25 000 元移到点⑨上来。这是申请专利权未被批准后进行的第一次决策。

点⑩：30 000(元)

点⑪：$(0.3\times 11\,000+0.4\times 8\,000+0.3\times 5\,000)\times 4=32\,000$(元)

点⑪的期望收益值大于点⑩的，所以应当选择自用方案，对出售方案进行修枝。把点⑩的 32 000 元移到点⑦上来。这是不申请专利权的第一次决策。

点⑥：$0.5\times 41\,200+0.5\times 25\,000-5\,000=28\,100$(元)

点⑦：32 000(元)

点⑦的期望收益值大于点⑥的期望收益值，所以应当选择不申请专利权方案。对申请专利权方案进行修枝。把点⑦的 32 000 元移到点④上来。这是第二次决策。

点②：$0.6\times 32\,000+0.4\times 0-15\,000=4\,200$(元)

点③：0

显然，应选择投资方案。因为投资方案可期望获利 4 200 元，而不投资则无收益。

(3) 进行决策。综合以上的决策分析，最优方案是投资发展新工序，如获成功，不申请专利权，自己使用而不出售的方案。

12.5 矩阵决策法

矩阵决策法是通过矩阵形式分析、选择决策的最优方案。它在经济活动中得到广泛的应用，尤其是对多种自然状态、多种方案的优选分析具有重要意义。

12.5.1 决策问题的矩阵结构形式

为了说明问题，我们假设 $a=\{a_1, a_2, \cdots, a_m\}$ 为决策者所有可能行动方案的

集合。如果把它看作一个向量，$a_i(i=1, 2, \cdots, m)$ 就是它的分量，可记作：$\boldsymbol{a}=(a_1, a_2, \cdots, a_m)$，称为方案向量。

又设 $\theta=\{\theta_1, \theta_2, \cdots, \theta_n\}$ 为所有自然状态的集合。如果也把它看作一个向量，$\theta_j(j=1, 2, \cdots, n)$ 就是它的分量，可记作：$\boldsymbol{\theta}=(\theta_1, \theta_2, \cdots, \theta_n)$ 称为自然状态向量。

若把状态 θ_j 发生的概率记作 $P(\theta_j)=P_j$，则 $\boldsymbol{P}=[P(\theta_1), P(\theta_2), \cdots, P(\theta_n)]$ 称为状态概率向量，全部状态概率之和应等于 1，即 $\sum_{j=1}^{n} P(\theta_j)=\sum_{j=1}^{n} P_j=1$。

当自然状态 θ_j，采取方案是 a_i 时，其相应的损益值为 $A(a_i, \theta_j)$，是 a_i 和 θ_j 的函数，简记为 a_{ij}。

即 $A(a_i, \theta_j)=a_{ij}$，而方案 a_i 的期望损益值为

$$E(a_i)=\sum_{j=1}^{n} P_j a_{ij} \quad i=(1, 2, \cdots, m)$$

现将自然状态、状态概率、行动方案、各方案对应的损益值和期望损益值用矩阵的形式列于表 12－11 中。

表 12－11 一般矩阵决策表

损益矩阵 \ 自然状态	θ_1	θ_2	$\cdots$	θ_n	期望损益值 $\boldsymbol{E}(a)$
行动方案 \ 状态概率	P_1	P_2	$\cdots$	P_n	
a_1	a_{11}	a_{12}	$\cdots$	a_{1n}	$E(a_1)$
a_2	a_{21}	a_{22}	$\cdots$	a_{2n}	$E(a_2)$
		……			
a_m	a_{m1}	a_{m2}	$\cdots$	a_{mm}	$E(a_m)$
决策→	$a_r=\max\limits_a[E(a_i)]$ 或 $a_x=\min\limits_a[E(a_i)]$				

表 12－11 中的损益矩阵又称为风险矩阵，如以 $\boldsymbol{B}$ 来代表，则

$$\boldsymbol{B}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \cdots\cdots & & \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix}$$

我们把 $E(a)$ 看作一个列向量或列矩阵，则

$$\boldsymbol{E}(a)=\begin{bmatrix}E(a_1)\\E(a_2)\\\vdots\\E(a_m)\end{bmatrix}$$

把状态概率向量 $\boldsymbol{P}$ 的转置矩阵记为

$$\boldsymbol{P}^{\mathrm{T}}=\begin{bmatrix}P_1\\P_2\\\vdots\\P_n\end{bmatrix}$$

显然，以上三者之间存在以下的关系：

$$\boldsymbol{E}(a)=\boldsymbol{B}\cdot\boldsymbol{P}^{\mathrm{T}}$$

因为

$$\boldsymbol{BP}^{\mathrm{T}}=\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\a_{21} & a_{22} & \cdots & a_{2n}\\ & \cdots\cdots & & \\a_{m1} & a_{m2} & \cdots & a_{mn}\end{bmatrix}\begin{bmatrix}P_1\\P_2\\\vdots\\P_n\end{bmatrix}=\begin{bmatrix}\sum_{j=1}^{n}P_j\,a_{1j}\\\sum_{j=1}^{n}P_j\,a_{2j}\\\vdots\\\sum_{j=1}^{n}P_j\,a_{mj}\end{bmatrix}=\begin{bmatrix}E(a_1)\\E(a_2)\\\vdots\\E(a_m)\end{bmatrix}=E(a)$$

当决策标准是收益时，应选择期望收益值最大的方案为最优方案，即 $a_r=\max\limits_{a}[E(a_i)]$。如果决策标准是损失时，则应选取其中期望损失值最小的方案为最优方案，即 $a_r=\min\limits_{a}[E(a_i)]$。

12.5.2 矩阵决策法的应用举例

【例 12－9】 某港机零配件代理商希望订购某最新的港机零配件。根据以往经验，新零配件的销售量可能为 50 个、100 个、150 个或 200 个。假定每个新零配件的订购价 4 元，销售价为 6 元，剩余新零配件的处理价为每个 2 元。又根据以往统计资料，新零配件销售量的规律如表 12－12 所示。

表 12－12 新零配件销售量的规律

销售量/个	50	100	150	200
概　率	0.2	0.4	0.3	0.1

试用矩阵决策法确定合理的订购数量。

【解】 (1) 编制损益矩阵,如表 12-13 所示。

表 12-13 编制损益矩阵

损益矩阵 \ 自然状态 概率 订购方案/个		市场销售状态 θ_j 50	θ_j 100	θ_j 150	θ_j 200	期望收益值/元 $\boldsymbol{E}(a)$
		0.2	0.4	0.3	0.1	
a_1	50	100	100	100	100	100
a_2	100	0	200	200	200	160
a_3	150	−100	100	300	300	140
a_4	200	−200	0	200	400	60

如方案 a_2 中的 a_{21} 的计算法为

$$a_{21} = 50 \times 2 - (100 - 50) \times 2 = 0(\text{元})$$

其他 $a_{ij}(i,\ j = 1,\ 2,\ 3,\ 4)$ 计算法类似,其结果如表 12-13 所示。

(2) 计算各方案的期望收益值。

$$\boldsymbol{E}(a) = \begin{bmatrix} E(a_1) \\ E(a_2) \\ E(a_3) \\ E(a_4) \end{bmatrix} = \boldsymbol{BP}^{\mathrm{T}} = \begin{bmatrix} 100 & 100 & 100 & 100 \\ 0 & 200 & 200 & 200 \\ -100 & 100 & 300 & 300 \\ -200 & 0 & 200 & 400 \end{bmatrix} \begin{bmatrix} 0.2 \\ 0.4 \\ 0.3 \\ 0.1 \end{bmatrix}$$

$$= \begin{bmatrix} 100 \\ 160 \\ 140 \\ 60 \end{bmatrix}$$

(3) 进行决策。

$$a_r = \max[E(a)] = \max(100,\ 160,\ 140,\ 60) = 160(\text{元})$$

计算表明,以 a_2 方案的期望收益值最大,因此决策选择 a_2 方案,即订购 100 个为最优方案。

如果决策的结果,有两个行动方案的期望值相同时,可进一步计算它们的全距和均方差。选取全距和均方差较小的方案为最优方案。

【例 12-10】 某厂对一问题进行决策,其有关资料如表 12-14 所示。试用矩

阵决策法选择最优方案。

表 12-14 矩阵决策表 单位：千元

收益矩阵 备选方案/本 \ 自然状态 状态概率	θ_1	θ_2	θ_3	θ_4	期望收益值/元 $E(a)$
	0.2	0.4	0.3	0.1	
a_1	40	50	60	70	55
a_2	20	40	60	90	53
a_3	50	70	30	50	56
a_4	30	50	60	80	56
a_5	30	50	50	50	46

【解】 (1) 计算各方案的期望收益值。

$$\boldsymbol{E}(a)=\begin{bmatrix}E(a_1)\\E(a_2)\\E(a_3)\\E(a_4)\\E(a_5)\end{bmatrix}=\boldsymbol{BP}^{\mathrm{T}}=\begin{bmatrix}40&50&60&70\\20&40&60&90\\50&70&30&50\\30&50&60&80\\30&50&50&50\end{bmatrix}\begin{bmatrix}0.2\\0.4\\0.1\\0.3\end{bmatrix}$$

$$=\begin{bmatrix}55\\53\\56\\56\\46\end{bmatrix}$$

$$a_r=\max(55,53,56,46)=56(\text{千元})$$

$$E(a_3)=56,\ E(a_4)=56$$

计算表明，a_3 和 a_4 两个方案的期望收益值同。

(2) 计算 a_3 和 a_4 两个方案的全距和均方差。

a_3 方案的全距 $=70-30=40$

a_4 方案的全距 $=80-30=50$

a_3 方案的均方差：

$$\sigma_3=[(50-56)^2\times0.2+(70-56)^2\times0.4+(30-56)^2\times0.1+(50-56)^2\times0.3]^{1/2}=12.806$$

a_4 方案的均方差：

$$\sigma_4 = [(30-56)^2 \times 0.2 + (80-56)^2 \times 0.4 + (60-56)^2 \times 0.1 + (80-56)^2 \times 0.3]^{1/2} = 18$$

(3) 进行决策。从以上的计算结果可以看出，a_3 方案的全距和均方差均小于 a_4 方案的全距和均方差，因此，最优方案为 a_3 方案。

12.6 敏感性分析

前面介绍的决策方法，其主要选优标准是期望损益值。但通常，自然状态概率及条件损益值是不容易估计准确的，从而期望损益值也就不十分准确。因此，有必要对状态概率或条件损益值数据的变动是否影响最优方案的选择进行分析，这种分析称为敏感性分析。我们希望最优方案对这些数据变动的反应是不敏感的，这样决策可靠性就大，决策错误的风险就较小。

【例 12-11】 某内河港口想从门座式起重机、桥式起重机两种岸边船舶装卸机械中购买一种。根据以往经验，如果今年件杂货船舶挂靠得多，购买门座式起重机可得利润 50 万元；购买桥式起重机要亏损 15 万元。如果今年集装箱船舶挂靠得多，购买门座式起重机会亏损 20 万元；而购买桥式起重机可得利润 100 万元。又根据该港过去件杂货和集装箱两货种的数据资料和目前经济发展趋势分析，该港口件杂货船舶挂靠多的概率是 0.7；相反，集装箱船舶挂靠多的概率为 0.3。问：该港口应该购买哪种岸边机械？

【解】 先列出状态概率和损益值，如表 12-15 所示。

表 12-15 不同港口机械的损益值 单位：万元

购买方案 \ 自然状态	件杂货船舶挂靠得多	集装箱船舶挂靠得多
状态概率	0.7	0.3
购买门座式起重机	50	−20
购买桥式起重机	−15	100

计算各方案的期望收益值：

购买门座式起重机：$0.7 \times 50 + 0.3 \times (-20) = 29$(万元)

购买桥式起重机：$0.7 \times (-15) + 0.3 \times 100 = 19.5$(万元)

显然，购买门座式起重机为最优方案。

如果最近件杂货船舶挂靠多的概率从 0.7 变到 0.8，情况会怎么样呢？

状态概率和损益值如表 12-16 所示。

表 12-16 状态概率和损益值 单位：万元

购买方案 \ 自然状态	件杂货船舶挂靠得多	集装箱船舶挂靠得多
状态概率	0.8	0.2
购买门座式起重机	50	−20
购买桥式起重机	−15	100

各方案的期望收益值为

$$购买门座式起重机：0.8 \times 50 + 0.2 \times (-20) = 36(万元)$$
$$购买桥式起重机：0.8 \times (-15) + 0.2 \times 100 = 8(万元)$$

此时，购买门座式起重机仍是最优方案，没有变化。再其次，如果最近件杂货船舶挂靠多的概率从 0.7 变到 0.6，情况又会怎么样呢？

状态概率和损益值如表 12-17 所示。

表 12-17 状态概率和损益值 单位：万元

购买方案 \ 自然状态	件杂货船舶挂靠得多	集装箱船舶挂靠得多
状态概率	0.6	0.4
购买门座式起重机	50	−20
购买桥式起重机	−15	100

各方案的期望收益值为：

$$购买门座式起重机：0.6 \times 50 + 0.4 \times (-20) = 22(万元)$$
$$购买桥式起重机：0.6 \times (-15) + 0.4 \times 100 = 31(万元)$$

现在，情况发生了变化，最优方案不再是购买门座式起重机，而是购买桥式起重机了。一个方案从最优方案转化为非最优方案，这是从一个量变到质变的过程。在这种转变过程中有一个概率值点。这个概率值点称为转折概率。最优方案的转化，都有转折概率。

设 P 代表件杂货船舶挂靠多的概率，$(1-P)$ 代表集装箱船舶挂靠多的概率。计算这两个方案的期望收益值，并令它们相等，可得

$$P \times 50 + (1-P) \times (-20) = P \times (-15) + (1-P) \times 100$$

简化后得

$$185P = 120$$

$$P = 0.65$$

这个 0.65 就是转折概率，当 $P > 0.65$ 时，购买门座式起重机是最优方案；当 $P < 0.65$ 时，购买桥式起重机是最优方案。

因此，在实际工作中，我们需要把概率值、损益值等在可能发生的范围内做几次不同的改动，并反复地进行计算。看看所得到的期望损益值是否相差很大，是否影响最优方案的选择。如果那些数据稍加变动，而最优方案保持不变，则这个方案是比较稳的，即灵敏度不高，决策可靠性大。反之，如果那些数据稍加变动，最优方案就从原来的变到另一个，则这个方案是不稳定的，即灵敏度高，决策可靠性小，需要进一步分析和研究改进措施。

12.7 效用概率决策方法

12.7.1 效用的含义

上述所讨论的决策方法，大多是以期望损益值作为决策标准的。但是，有时这样做既不合理，也不符合实际。

决策是由决策人做出的，决策人的经验、才智、胆识和判断能力等主观因素，不能不对决策产生重要影响。如果完全以期望值的大小作为决策标准，就会把决策过程变成为机械地计算期望损益值的过程，从而把决策人的主观作用排除在外，这当然是不够合理的。此外，决策还受到一些特殊情况的影响。例如，有两个方案：方案甲以 0.5 的概率获利 4 000 元，以 0.5 的概率蒙受损失 2 000 元；方案乙以 100%的概率获利 500 元，试做出最优方案的抉择。尽管方案甲的期望利润值 $= 0.5\times4\,000+0.5\times(-2\,000)=1\,000$(元)，是方案乙获利 500 元的两倍，但对相当多的人来说，却宁愿选择方案乙，因为选择方案乙肯定盈利 500 元，并且可避免冒 50%的损失 2 000 元的风险。当然也有人会选择方案甲，企图盈利 1 000 元。

这就涉及决策人对风险的态度了，这种态度对决策起了重要的作用。一般来说，当同一决策要重复多次或风险损失数值较小时，决策人的兴趣会与期望损益值的高低大体一致；当同一决策只进行一次，且包含较大风险时，决策人的兴趣往往会与期望损益值之间发生较大差异。

对于相同的期望损益值，不同决策人的反应也不一定相同。这是由于决策人的个人性格和处理条件不同所形成的。前者如保守型决策人与进攻型决策人的反应会不相同，后者如大企业的经理对上万元的风险损失看作是一桩小事，而小企业的经理则会看作是与企业兴衰攸关的大事。即使是同一决策人，由于时间和条件

不同，对于相同机会的反应也不相同。这种决策人对于期望收益和损失的独特兴趣、感受和取舍反应，称为效用。效用代表着决策人对于风险的态度，也是决策人胆略的一种反映。

12.7.2 效用曲线

决策者对具有不同风险的相同期望损益值，会给出不同的效用值。若用横坐标代表损益值，用纵坐标代表效用值，把决策者对风险态度的变化关系绘出一条曲线，就称为决策人的效用曲线。效用可以通过计算效用值和绘制效用曲线的方法来衡量。

假定有一个方案甲，以 0.5 的概率得收益 200 元，以 0.5 的概率损失 100 元；另有一个方案乙，以 1.0 的概率得收益 25 元，如图 12-6 所示。

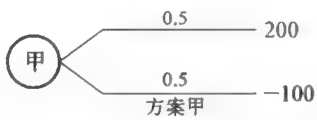


图 12-6　两个可供选择的方案

现在考虑一下，这两个方案究竟应选哪个？可能大多数人会选择方案乙，尽管方案甲的期望值是 $0.5\times200+0.5\times(-100)=50$(元)，而方案乙的期望值只有 $1.0\times25=25$(元)。决策人之所以宁愿选择 25 元的方案乙，是因为他不愿意负担遭受损失的风险。

首先，我们规定，效用值用概率形式表现，是介于 0 与 1 之间的数值。我们以损益最大值(本例为 200 元)的效用值作为 1，以损益最小值(本例为 −100 元)的效用值作为 0，如图 12-7(a)所示。

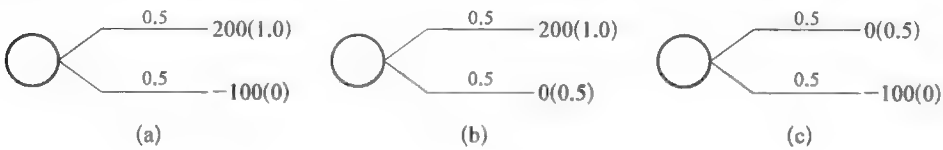


图 12-7　效用值的确定

已经知道，决策人宁愿接受肯定的 25 元，而不愿意接受方案甲(见图 12-6)，这说明 25 元的效用大于方案甲的效用。假定再问，方案乙的收益改为 10 元怎么样？如果决策人宁愿接受肯定的 10 元，则说明 10 元的效用还是大于方案甲。再问，如果方案乙的收益改为损失 10 元怎么样？现在决策人可能宁愿接受方案甲而不愿付出 10 元，这说明这时方案甲的效用大于 −10 元的效用。经过几次询问，不断地改变肯定得到的损益值，最后总可以使决策人感到某一损益值与方案甲具有

同样的效用，此时，决策者对两种方案具有同样的兴趣。设最后判断的损益值为0元，则与0元相应的效用值与方案甲的期望效用值 $0.5\times1.0+0.5\times0=0.5$ 相等，即损益值0元的效用值是0.5。

现在，以损益值为横坐标，以效用值为纵坐标，就可以画出效用曲线，如图12-8所示。对应于损益值0和效用值0.5，就得到效用曲线上的一个点 A。

其次，以0.5的概率得收益值200元、0.5的概率得收益值0元作为方案，如图12-7(b)所示。重复上述过程，假定经过多次询问，最后判断80元的效用与这个方案的效用相等，则相当于80元的效用值为

$$0.5\times1.0+0.5\times0.5=0.75$$

这样，在0与200元之间又得到对应于损益值80元和效用值0.75的一个点 B。

再次，以0.5的概率得收益值0元、0.5的概率得收益值-100元作为方案，如图12-7(c)所示。重复上述过程，假定经过几次询问，最后判断出来决策人对-60元的效用等于这个方案的效用，则相当于-60元的效用值为

$$0.5\times0.5+0.5\times0=0.25$$

这样，又可得到图12-8对应于损益值-60元和效用值0.25的一个点 C。

采用同样办法，可以得到许多这样的点，把它们连接起来，就成为所谓的效用曲线(见图12-8)。从这条线上可以找出对应于各个损益值的效用值，反过来，也可以找出对应于各个效用值的损益值。

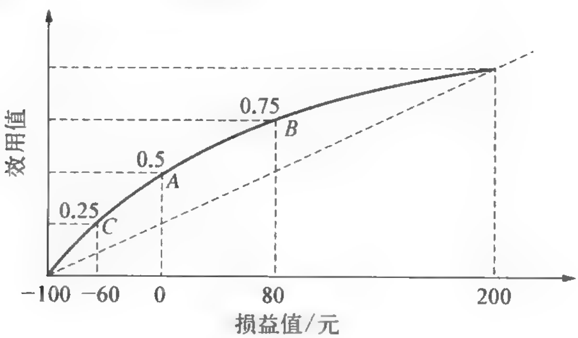


图12-8 效用曲线

12.7.3 效用曲线的类型

效用曲线有三种类型(见图12-9)：

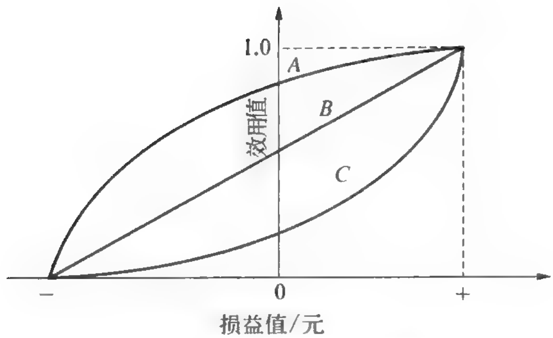


图 12-9 效用曲线的类型

(1) 沿用上例，曲线 A 代表决策人宁愿选择肯定得到 50 元的方案，而不愿选择 $0.5\times200+0.5\times(-100)=50$（元）的方案。这种类型的决策人对于利益反应比较迟缓，而对损失比较敏感，是一种不求大利、避免风险、谨慎小心的保守型决策人。

(2) 曲线 C 代表决策人宁愿选择 $0.5\times200+0.5\times(-100)=50$（元）的方案，而不愿接受得到 50 元的方案。这种决策人对于损失反应迟缓，而对利益比较敏感，是一种谋求大利、不怕风险的进取型决策人。

(3) 曲线 B 所代表的是一种中间型的决策人。他认为肯定得到 50 元的方案和 $0.5\times200+0.5\times(-100)=50$（元）的方案没有差别，有相等的效用值。也就是说，他只要利用期望损益值作为选择方案的标准就可以了，而不需要利用效用曲线。

通过对大量事实的研究证明，大多数决策者属于保守型，而其他两类决策者属于少数。

12.7.4 效用曲线的应用

现在通过一个例子说明效用曲线的应用方法。

【例 12-12】 我们重新考虑图 12-2。根据原来的计算，建设大泊位方案的期望收益值为 680 万元，建设小泊位方案的期望收益值为 460 万元，按期望值标准判断前者为最优方案。

现在我们首先画出决策树，并把各种方案的损益值标写在各个概率枝的末端（见图 12-10）。这些损益值是根据图 12-2 的数据，按照 10 年累计（扣除投资额后）计算出来的，如图 12-10 所示。

其次，绘出决策人的效用曲线。我们以最大损益值 1 400 万元的效用值作为 1.0，以最小损益值 −1 000 万元的效用值作为 0，向这个问题的决策人提出一系列

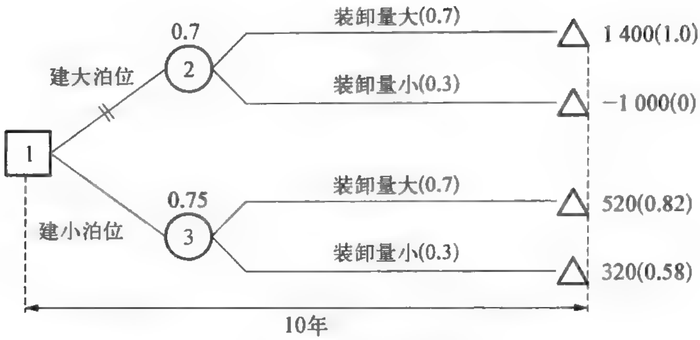


图 12 - 10 用效用值评价的决策树

问题，找出对应于若干个损益值的效用值，这样就可以画出效用曲线，如图 12 - 11 所示。

画出了效用曲线后，就可以反过来找出对应于原决策问题各个损益值的效用值：520 万元的效用值等于 0.82，320 万元的效用值等于 0.58。把对应的效用值写在图 12 - 10 中各损益值之后。现在可以用效用值代替原损益值来计算每一方案的效用期望值，效用曲线如图 12 - 11 所示。

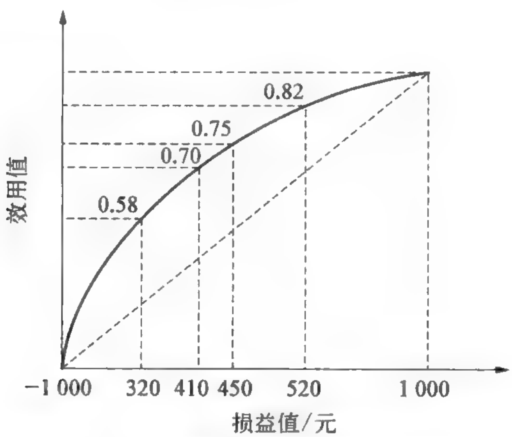


图 12 - 11 建设工厂规模问题的效用曲线

建设大工厂方案的效用期望值是：

$$0.7 \times 1.0 + 0.3 \times 0 = 0.70$$

建设小工厂方案的效用期望值是：

$$0.7 \times 0.82 + 0.3 \times 0.58 = 0.75$$

由此可见，如果以效用期望值作为评价标准，小工厂方案反而较好。为什么会出现这种情况呢？这是因为决策人是保守型的，他不想冒太大的风险。从曲线上可以测出，效用值 0.70 只相当于损益值 410 万元，这大大低于原来的期望值 680 万元；效用值 0.75 也只相当于损益值 450 万元，这也小于原来的期望值 460 万元。

思考与练习

1. 某杂志零售商店对《大众电影》每月需求量份数根据历史资料估计如题表 12－1 所示。

题表 12－1

每月销售量/千本	1	2	3	4	5
概　率	0.10	0.15	0.30	0.25	0.20

这种杂志每本零售价为 0.8 元，每本批进价为 0.6 元。试问该店做出购进杂志数量的最优决策：

(1) 若卖不出去的杂志可以按批进价退回，该店每月应订购多少本？

(2) 若卖不出去的杂志不能退回，但下月可以按零售价对折，即按每本 0.4 元出售，这样，该店每月应订购多少本？

2. 某钟表公司计划通过它的销售网推销一种低价钟表，计划零售价为每块 10 元。对这种钟表有三个设计方案：第一方案需一次投资 10 万元，投产后每块成本 5 元；第二方案需一次投资 16 万元，投产后每块成本 4 元；第三方案需一次投资 25 万元，投产后每块成本 3 元。这种钟表需求量不确切，但估计有 3 种可能：

E_1 为 30 000 块；E_2 为 120 000 块；E_3 为 200 000 块；并预测这 3 种需求量的概率如题表 12－2 所示。

题表 12－2

需求量	E_1	E_2	E_3
概　率	0.15	0.75	0.10

要求：

(1) 分别用期望收益决策法和期望损失决策法决定该公司应采用哪一个设计方案？

(2) 如有一个部门能帮助调查市场的确切需要量，该公司最多愿意花多少调查费用？

(3) 进行灵敏性分析，确定用期望收益决策法决策时的转折概率。

3. 某副食品商场在夏季购进新鲜草莓，每公斤进价 2 元，售价 4 元，每公斤销售可获利 2 元；如果当日不能售出，由于变质，每公斤只能收回 1 元。现市场需求情况不清楚，但有上年同期销售资料，如题表 12－3 所示。

题表 12－3

日销售量/kg	完成日销售量的天数
400	12
450	24
500	38
550	16
$\sum$	90

试用边际决策法为该商场做出最佳进货量。

4. 某镜片厂生产一种名为广角摄影镜头的新产品，市场价格为每件 200 元。生产方案有 3 种：一是自制设备生产；二是租用设备生产；三是与外商合资生产。经过计算不同经营方式（不同方案）需要的固定成本和可变成本，如题表 12－4 所示。

题表 12－4

	制设备	租用设备	与外商合资
固定资产总额(元)	1 200 000	400 000	640 000
每件可变成本(元)	60	100	80

市场销售预测有下面 3 种可能：一是畅销（年销售 30 000 件）的概率为 0.2；二是中等销售（年销售 20 000 件）的概率为 0.7；三是滞销（年销售 10 000 件）的概率为 0.1。要求：

(1) 计算编制一张广角镜头经营决策条件损益表。

(2) 利用决策树法为该厂广角镜头经营方式做出最优方案。

5. 某工程队承担一座桥梁的施工任务。由于施工地区夏季多雨，需停工 3 个月。在停工期间，该工程队可将施工机械搬走或留在原处。如搬走，需搬运费 1 800元；如留原处，一种方案是花 500 元筑一护堤，防止河水涨发生高水位的侵袭。若不筑护堤，发生高水位侵袭时将损失 1 000 元。如下暴雨发生洪水时，则不管是否筑护堤，施工机械留在原处将受到 60 000 元的损失。根据历史资料，该地区夏季出现洪水的概率为 2%，出现高水位的概率为 25%，出现正常水位的概率为

73%。试用决策树法为该施工队做出最优决策。

6. 某企业从事石油钻探工作，它准备与某石油公司签订合同，钻探一片可能产油的勘探点。该企业可供选择的方案有两种：一是先做地震实验，看实验结果如何，再决定是否要钻井；二是不做地震实验，只凭经验决定是否要钻井。做实验要花 3 000 元，钻井要花 10 000 元。如钻出石油可获得 50 000(即公司付给企业 50 000)元；若钻不出石油，则企业没有收入。

根据历史资料的分析估计：做地震实验，其结果良好的概率为 0.6，不好的概率为 0.4；经地震实验为良好时，钻井出油的概率为 0.8，钻井不出油的概率为 0.2；经地震实验为不好时，钻井出油的概率为 0.1，钻井不出油的概率为 0.9；不经地震实验而钻井时，出油的概率为 0.55，不出油的概率为 0.45. 试用决策树法为该企业做出最优方案。

7. 某商场经销某种商品，进货成本为每公斤 3 元，销售价格为每公斤 4 元。如果在一周内不能出售，由于质量降低和部分霉烂等原因要降价出售，平均每公斤只能回收 2 元。根据历史资料分析估计，这种商品的每周销售概率如题表 12 - 5 所示：

题表 12 - 5

每周销售/kg	概　　率
10 000	0.08
11 000	0.12
12 000	0.30
13 000	0.40
14 000	0.10

试用矩阵决策法为该商场做出最优进货方案。

13 贝叶斯决策

在上一章中，我们研究了风险型决策的基本原理和几种常用的方法。风险型决策就是在自然状态不确定的情况下进行决策的，它的基本方法是将状态变量视为随机变量，将决策者对状态变量发生的可能性的估计用概率分布的形式表示，并在决策分析中使用。

作为决策者，如何对自然状态发生的可能性做出估计？如何利用新取得的各种信息情报对以往的估计加以修正？由于对自然状态发生的可能性的估计的正确与否直接决定着我们所做出的决策的优劣，因此本章介绍的方法就是要使决策者通过有效的方法对各种自然状态发生的可能性做出更精确的估计，从而得出更为优良的决策方案。

本章将介绍两种估计自然状态发生可能性的方法，即利用经验情报做估计和利用抽样情报做估计；并介绍一种修正方法，即贝叶斯(Bayes)公式修正法。

实际上，本章的决策方法是通过对决策的问题进行更多的调查了解后再做出决策，这样做法的基本优点是能提高决策的正确性，但也会存在一些缺点：如只有在调查完毕并分析出结果后，才进行决策分析，可能会延误做决策的时间；另外，进行调查通常需要耗费人力、物力和财力，从而使决策成本提高。所以，在掌握了各种决策原理和方法后，将它们用于各种具体问题时，应当权衡其利弊，使所做出的决策能尽量及时、经济和有优良效果。

13.1 先验概率分布

做决策分析时，最先确定的各种自然状态的概率一般称为先验概率分布，它是在做任何实验或调查以前就确定了的。若根据试验或调查所获得的情报，对先前确定的先验概率分布加以修正，而得到关于自然状态的新的概率分布，则称之为后验分布。

13.1.1 客观的先验概率分布

一般决策问题的先验概率分布，可由经验上已获得的情报或资料来推测。例

如，一个码头的某种货物在过去 30 天内的出口记录如表 13-1 所示。

表 13-1　港口出口记录

出口量/万吨	天　　数	频　　率
10 以下	3	0.1
10～30	9	0.3
30～50	15	0.5
50 以上	3	0.1

由这些资料，我们可以确定未来任何一天的出口量(即自然状态)的概率分布如表 13-2 所示。

表 13-2　港口出口概率分布

出口量/万吨	10 以下	10～30	30～50	50 以上
概　率	0.1	0.3	0.5	0.1

又如，我们可以用某一段时期内每批出口产品所包含的不合格品数目来估计该产品不合格品率的概率分布，用过去历年秋季广州市火灾的次数来估计明年秋季火灾次数的概率分布，等等。这些都是先验分布例子。对这些自然状态的先验概率的估计是根据某些客观的情报或证据得出的，故称其为客观先验分布。

13.1.2　主观的先验概率分布

如果缺乏有关自然状态的客观上的情报，先验概率分布必须主观地指定，就得到所谓主观的先验概率分布。这种概率分布往往并非单纯由猜测得出的，而常常由决策者小心分析自然状态的各种情况、评估各种自然状态出现的可能性大小之后再予以指定。很可能决策者拥有一些关于自然状态的经验、知识或情报等，尽管这些并不以客观的或计量化的方式存在，但它们对于指定先验概率分布却是有帮助的。

我们举例来说明决策者如何指定主观的先验概率分布。假定有个决策问题，其自然状态为未来 6 个月内的 BDI 指数变化，设有两种可能的自然状态，分别用 S_1 和 S_2 表示，其中 S_1 为 BDI 指数将下降；S_2 为 BDI 指数将维持不变或上升。

如果能获得此自然状态的概率分布，即 S_1 和 S_2 出现的概率，就可以利用它做出可靠的决策。假如决策者已有从事大宗散货研究多年的经历，对于预测未来的航运市场形式有丰富的经验，那么他可以运用过去观察到的经济状况、全球政治形势、航运市场景气程度等因素与 BDI 指数之间的关系去推测 S_1 和 S_2 发生的概率，

而不仅仅是凭预感或直觉去指定这些概率。如果决策者不满足于单凭他自己拥有的知识去指定先验概率,则他还可以再请教一些航运专家、经济学者、市场研究机构等,综合他们的意见后再来指定这些先验概率。

主观的先验概率分布,反映决策者对未知的自然状态的不确定性意见。这种概率分布有时并不需要有经验上的证据来支持,两个决策者处理同一个决策问题,在完全相同的情况下指定的先验概率分布却很可能不一致。一旦先验概率分布被指定,不论它是主观的或是客观的,都可当作一般的概率分布,用各种概率运算法则进行运算。

13.2 贝叶斯定理与后验概率分布

决策者为了对决策问题的自然状态有更多的了解而进行统计调查,我们称通过调查而获得的信息为补充信息,利用贝叶斯定理将补充信息和先验分布结合起来,便产生了一种综合信息,即为后验概率分布。后验概率分布可用来做出较为正确的决策,本节将介绍贝叶斯定理及其应用。

设自然状态 θ 有 k 种,分别用 $\theta_1,\ \theta_2,\ \cdots,\ \theta_k$ 表示,$P(\theta_i)$表示自然状态 θ_i 发生的先验概率分布,用 x 表示调查结果,$P(x|\theta_i)$表示在状态 θ_i 条件下调查结果刚好为 x 的概率。通过调查得到结果 x,这样的结果包含有关于自然状态 θ 的信息,利用这些信息可对自然状态 $\theta_i(i=1,\ 2,\ \cdots,\ k)$ 发生的概率重新认识,并加以修正,修正后的概率为

$$P(\theta_i \mid x)=\frac{P(x \mid \theta_i)P(\theta_i)}{\sum_{j=1}^{k}P(x \mid \theta_j)P(\theta_j)} \quad (i=1,\ 2,\ \cdots,\ k)$$

这就是贝叶斯公式。一般来讲,这时对各种自然状态 $\theta_1,\ \theta_2,\ \cdots,\ \theta_k$ 发生的概率做出的估计 $P(\theta_1|x),\ P(\theta_2|x),\ \cdots,\ P(\theta_k|x)$ 比先验概率分布更为准确。我们称 $P(\theta_i|x)$为 θ_i 发生的后验概率。

【例 13-1】 某自动生产设备在生产过程中可能正常亦可能不正常,正常时产品的合格率为 80%,不正常时产品的合格率为 30%。从某时刻生产的产品中抽取一件进行检验,要求我们根据这件产品的情况来判断设备是否正常。

该问题的自然状态有两种,即设备正常和设备不正常,分别用 θ_1 和 θ_2 表示,假设我们对该设备以往的生产情况一无所知,那么判断设备是否正常的可能性相等,即先验概率为

$$P(\theta_1)=0.5,\ P(\theta_2)=0.5$$

由于两者的概率相等,实际上无法判断出设备究竟是否正常。但如果我们从某时刻的产品中抽取一件产品,若发现为合格品,即抽样的结果为 $x=$“合格品”,这就得到了一种补充的信息,容易算出

$$P(\text{合格品} \mid \theta_1) = 0.8$$
$$P(\text{合格品} \mid \theta_2) = 0.3$$

利用贝叶斯公式得

$$\begin{aligned} P(\theta_1 \mid \text{合格品}) &= \frac{P(\text{合格品} \mid \theta_1)P(\theta_1)}{P(\text{合格品} \mid \theta_1)P(\theta_1) + P(\text{合格品} \mid \theta_2)P(\theta_2)} \\ &= \frac{0.8 \times 0.5}{0.8 \times 0.5 + 0.3 \times 0.5} = 0.73 \\ P(\theta_2 \mid \text{合格品}) &= \frac{P(\text{合格品} \mid \theta_2)P(\theta_2)}{P(\text{合格品} \mid \theta_1)P(\theta_1) + P(\text{合格品} \mid \theta_2)P(\theta_2)} \\ &= \frac{0.3 \times 0.5}{0.8 \times 0.5 + 0.3 \times 0.5} = 0.27 \end{aligned}$$

即抽得一件产品为合格品后算得设备为正常的概率是 0.73,设备不正常的概率是 0.27,故应判断此时设备正常,即 $\theta = \theta_1$。

若从某时刻生产的产品中抽取到的一件产品为不合格品,同样利用贝叶斯公式算得

$$P(\theta_1 \mid \text{不合格品}) = 0.22,\ P(\theta_2 \mid \text{不合格品}) = 1 - 0.22 = 0.78$$

故应判断此时设备不正常,即 $\theta = \theta_2$。

现在将情况改变一下,如果从某时刻生产的产品中连续抽取两件产品,并检查它们是否合格,然后再判断设备此时是否正常。若抽样的结果为 $x=$“合·合”,即两件产品皆为合格品,容易算得

$$P(\text{合} \cdot \text{合} \mid \theta_1) = P(\text{合} \mid \theta_1) \cdot P(\text{合} \mid \theta_1) = 0.8 \times 0.8 = 0.64$$
$$P(\text{合} \cdot \text{合} \mid \theta_2) = P(\text{合} \mid \theta_2) \cdot P(\text{合} \mid \theta_2) = 0.3 \times 0.3 = 0.09$$

由贝叶斯定理得

$$\begin{aligned} P(\theta_1 \mid \text{合} \cdot \text{合}) &= \frac{P(\text{合} \cdot \text{合} \mid \theta_1)P(\theta_1)}{P(\text{合} \cdot \text{合} \mid \theta_1)P(\theta_1) + P(\text{合} \cdot \text{合} \mid \theta_2)P(\theta_2)} \\ &= \frac{0.64 \times 0.5}{0.64 \times 0.5 + 0.09 \times 0.5} = 0.877 \end{aligned}$$

利用概率的性质得

$$P(\theta_2 \mid 合\cdot合) = 1 - P(\theta_1 \mid 合\cdot合) = 0.123$$

据上面求出的后验概率，我们应判断此时设备为正常，即 $\theta = \theta_1$。

若抽样的结果为 x="合・不"，即抽得的第一件产品为合格品，第二件产品为不合格品，容易算得

$$P(合\cdot不 \mid \theta_1) = P(合 \mid \theta_1)\cdot P(不 \mid \theta_1) = 0.8 \times 0.2 = 0.16$$
$$P(合\cdot不 \mid \theta_2) = P(合 \mid \theta_2)\cdot P(不 \mid \theta_2) = 0.3 \times 0.7 = 0.21$$

由贝叶斯公式以及概率的性质知：

$$\begin{aligned} P(\theta_1 \mid 合\cdot不) &= \frac{P(合\cdot不 \mid \theta_1)P(\theta_1)}{P(合\cdot不 \mid \theta_1)P(\theta_1) + P(合\cdot不 \mid \theta_2)P(\theta_2)} \\ &= \frac{0.16 \times 0.5}{0.16 \times 0.5 + 0.21 \times 0.5} = 0.432 \end{aligned}$$

$$P(\theta_2 \mid 合\cdot不) = 1 - P(\theta_1 \mid 合\cdot不) = 0.568$$

因此，应判断此时设备不正常。

同样的方法，可以求出抽出的两件产品皆为不合格品，即 x="不・不"，或第一件产品为不合格品，第二件产品为合格品，即 x="不・合"的后验概率。

$$P(\theta_1 \mid 不\cdot不) = 0.075$$
$$P(\theta_2 \mid 不\cdot不) = 0.925$$
$$P(\theta_1 \mid 不\cdot合) = 0.432$$
$$P(\theta_1 \mid 不\cdot合) = 0.568$$

根据这些后验概率，合理的判断应当是：若两件产品皆为不合格品，判断设备此时不正常；或第一件为不合格品，第二件产品为合格品，判断设备此时不正常。

很多情况下，我们容易知道某一事件或试验结果(x)在各种状态下发生的概率 $P(x|\theta_i)$，因此上面的贝叶斯公式很有实用价值。在实际问题中，我们还会遇到这样的问题，我们虽不知道状态变量究竟是什么(只知道 $\theta = \theta_1$ 的概率为 $P(\theta_1)$)，但我们却想知道某一事件或试验结果(x)出现的概率 $P(x)$，这时我们仍可利用 $P(x|\theta_i)$来计算，具体计算公式为

$$P(x) = \sum_{i=1}^{k} P(x \mid \theta_i)P(\theta_i)$$

$P(x)$即为 x 在各种状态下可能出现的概率的综合值，故该公式被称为全概率公式。

【例 13－2】 一码头进口某种商品，其货源可能来自甲、乙、丙三个出口商之一

（假定其商标与价格都一样）。按以往的习惯，该码头有20%的时间进口甲商的产品，有40%的时间进口乙商的产品，有40%的时间进口丙商的产品。经过调查，甲、乙、丙三出口商产品的合格率分别是95%、92%和90%，那么：

(1) 这个码头进口的这种商品的总合格率为多少？

(2) 如果某人买一件这种商品，发现其为不合格品，我们应如何推断此时码头正在进口哪个出口商的产品？

用x表示商品为合格品，则$P(x)$即为这个码头进口的产品的合格率，此处状态变量θ为甲、乙或丙。由题意知：

$$P(\text{甲})=0.2 \quad P(\text{乙})=0.4 \quad P(\text{丙})=0.4$$

$$P(x\mid\text{甲})=0.95 \quad P(x\mid\text{乙})=0.92 \quad P(x\mid\text{丙})=0.90$$

有全概率公式：

$$P(x)=0.95\times0.2+0.92\times0.4+0.90\times0.4=0.918$$

若用$\bar{x}$表示“某商品为不合格品”，容易知道：

$$P(\bar{x}\mid\text{甲})=0.05 \quad P(\bar{x}\mid\text{乙})=0.08 \quad P(\bar{x}\mid\text{丙})=0.10$$

因此由贝叶斯公式知：

$$\begin{aligned}P(\text{甲}\mid\bar{x})&=\frac{P(\bar{x}\mid\text{甲})P(\text{甲})}{P(\bar{x}\mid\text{甲})P(\text{甲})+P(\bar{x}\mid\text{乙})P(\text{乙})+P(\bar{x}\mid\text{丙})P(\text{丙})}\\&=0.122\end{aligned}$$

同样可计算：

$$P(\text{乙}\mid\bar{x})=0.390 \quad P(\text{丙}\mid\bar{x})=0.488$$

因此，码头此时最有可能在进口丙出口商的产品。

13.3 后验决策及其优良性

建立在先验概率分布的基础上而做出的决策称为先验决策，利用后验概率分布做出的决策称为后验决策。理论上已证明，任何后验补充情报信息都不会给决策带来坏处。由于后验概率分布正是由先验分布经过补充一些情报后而产生的，故基于后验概率分布而做出的后验决策总是优于先验决策。

假定某一决策问题的自然状态θ为θ_1，θ_2，…，θ_n；它们的先验概率分布为$P(\theta=\theta_1)$，$P(\theta=\theta_2)$，…，$P(\theta=\theta_n)$；可以采取的行动a为a_1，a_2，…，a_m；损失

函数为 $R(\theta, a)$；令补充情报值为 x，在上节所述的例子中，我们通过抽出一件产品来补充情报信息，若抽出的产品为合格品，则 $x=$“合”；若抽出两件产品来补充情报信息，则 x 可能等于“合·合”，“不·不”，“合·不”，或“不·合”。根据情报值 x，我们采取某个行动 $\delta(x)$，$\delta(x)$ 可能为 a_1，a_2，…，或 a_m，称 $\delta(x)$ 为一个决策方案。

对某一决策方案 $\delta(x)$，在任一状态 $\theta=\theta_i$ 下，当情报值 x 确定后，它所对应的行动 $\delta(x)$ 也确定了，从而行动 $\delta(x)$ 的损失值 $R(\theta_i, \delta(x))$ 也就随之确定了。对于一个好的决策方案，应要求 $R(\theta_i, \delta(x))$ 较小，但是评价一个决策方案的好坏不能只看一次情报所取的值，应当用各种情报下的平均效果来衡量。因此，在状态 θ_i 下，决策方案的好坏应以 $R(\theta_i, \delta(x))$ 对情报值 x 的数学期望的大小为标准，即

$$P(\theta_i, \delta)=E_{x|\theta=\theta_i}[R(\theta_i, \delta(x))]$$

称 $P(\theta_i, \delta)$ 为在状态 θ_i 下决策方案 $\delta(x)$ 的风险值。风险值表示在固定状态 δ_i 下，当出现各种不同情报值时按决策方案采取行动的平均损失。

在不同状态下，同一决策方案的风险值不一样，一个决策方案的好坏应综合反映其在各种不同状态下的风险值。即我们应当用：

$$B(\delta)=E_\theta[P(\theta\delta)]=\sum_{i=1}^{n}P(\theta_i, \delta)P(\theta=\theta_i)$$

来衡量一个决策方案的好坏。称 $B(\delta)$ 为决策方案 $\delta(x)$ 的贝叶斯风险，它反映这一决策方案的平均损失。

【例 13-3】 在上节所述的例子中，如果设备正常，而判断为不正常，会损失 1 500元；判断为正常，损失为 0。若设备不正常，而判断为正常会损失 2 000 元，判断为不正常则损失为 0。我们来求各种决策方案的风险值和贝叶斯风险。

用 a_1 表示“判断设备正常”，a_2 表示“判断设备不正常”，该决策问题的损失矩阵如表 13-3 所示。

表 13-3 损失矩阵表

θ	$P(\theta=\theta_i)$	a_1	a_2
θ_1	1/2	0	1 500
θ_2	1/2	2 000	0

先研究抽取一件产品进行检验的情况。这时共有 4 个决策方案可供选择：

$$\delta_1(x)\begin{cases}a_1 & x=\text{“合”}\\ a_2 & x=\text{“不”}\end{cases},\quad \delta_2(x)\begin{cases}a_1 & x=\text{“不”}\\ a_2 & x=\text{“合”}\end{cases},$$

$$\delta_3(x)\begin{cases}a_1 & x=\text{“合”}\\a_1 & x=\text{“不”}\end{cases},\quad \delta_4(x)\begin{cases}a_2 & x=\text{“合”}\\a_2 & x=\text{“不”}\end{cases}$$

我们知道：

$$P(\theta=\theta_1)=\frac{1}{2},\ P(\theta=\theta_2)=\frac{1}{2}$$

$$P(x=\text{“合”}\mid\theta=\theta_1)=0.8,\ P(x=\text{“不”}\mid\theta=\theta_1)=0.2$$

$$P(x=\text{“合”}\mid\theta=\theta_2)=0.3,\ P(x=\text{“不”}\mid\theta=\theta_2)=0.7$$

对于决策方案 $\delta_1(x)$：

$$R(\theta_1,\delta_1(\text{合}))=R(\theta_1,a_1)=0$$

$$R(\theta_1,\delta_1(\text{不}))=R(\theta_1,a_2)=1\,500$$

$$R(\theta_2,\delta_1(\text{合}))=R(\theta_2,a_1)=2\,000$$

$$R(\theta_2,\delta_1(\text{不}))=R(\theta_2,a_2)=0$$

于是 $\delta_1(x)$的风险值为

$$\begin{aligned}P(\theta_1,\delta_1)&=E_{x|\theta=\theta_1}[R(\theta_1,\delta_1(x))]\\&=R(\theta_1,\delta_1(\text{合}))P(x=\text{“合”}\mid\theta=\theta_1)+\\&\quad R(\theta_1,\delta_1(\text{不}))P(x=\text{“不”}\mid\theta=\theta_1)\\&=0\times0.8+1\,500\times0.2=300(\text{元})\\P(\theta_2,\delta_1)&=E_{x|\theta=\theta_2}[R(\theta_2,\delta_1(x))]\\&=R(\theta_2,\delta_1(\text{合}))P(x=\text{“合”}\mid\theta=\theta_2)+\\&\quad R(\theta_2,\delta_1(\text{不}))P(x=\text{“不”}\mid\theta=\theta_2)\\&=2\,000\times0.3+0\times0.7=600(\text{元})\end{aligned}$$

故决策方案 $a_1(x)$贝叶斯风险为

$$\begin{aligned}B(\delta_1)&=P(\theta_1,\delta_1)P(\theta=\theta_1)+P(\theta_2,\delta_1)P(\theta=\theta_2)\\&=300\times\frac{1}{2}+600\times\frac{1}{2}=450(\text{元})\end{aligned}$$

用同样的方法，我们可以求得决策方案 $\delta_2(x)$的贝叶斯风险为

$$B(\delta_2)=1\,300(\text{元})$$

对于决策方案 $\delta_3(x)$：

$$R(\theta_1,\delta_3(\text{合}))=R(\theta_1,a_1)=0$$

$$R(\theta_1, \delta_3(\text{不})) = R(\theta_1, a_1) = 0$$
$$R(\theta_2, \delta_3(\text{合})) = R(\theta_2, a_1) = 2\,000$$
$$R(\theta_2, \delta_3(\text{不})) = R(\theta_2, a_1) = 2\,000$$

于是 $\delta_3(x)$的风险值为

$$P(\theta_1, \delta_3) = E_{x|\theta=\theta_1}[R(\theta_1, \delta_3(x))] = 0\times 0.8 + 0\times 0.2 = 0(\text{元})$$
$$P(\theta_2, \delta_3) = E_{x|\theta=\theta_2}[R(\theta_2, \delta_3(x))] = 2\,000\times 0.3 + 2\,000\times 0.7$$
$$= 2\,000(\text{元})$$

其贝叶斯风险为

$$B(\delta_3) = P(\theta_1, \delta_3)P(\theta=\theta_1) + P(\theta_2, \delta_3)P(\theta=\theta_2)$$
$$= 0\times\frac{1}{2} + 2\,000\times\frac{1}{2} = 1\,000(\text{元})$$

用同样的方法可求得决策方案 $\delta_4(x)$的贝叶斯风险为

$$B(\delta_4) = 750(\text{元})$$

因此，我们若用贝叶斯风险衡量，方案 $\delta_1(x)$优于其他 3 种决策方案。

若抽取两件产品来补充情报信息，这时决策方案共有 8 个，分别记为 δ_1', δ_2', δ_3', δ_4', δ_5', δ_6', δ_7', δ_8'，各个决策方案的风险值和贝叶斯风险如表 13－4 所示。

表 13－4　抽取两件产品时各个决策方案的风险值和贝叶斯风险

状态 θ 先验分布 $P(\theta)$	$\theta=\theta_1$ 0.5			$\theta=\theta_2$ 0.5		
$P(x\mid\theta=\theta_1)$	合·合 0.64	1合·1不 0.32	不·不 0.04	合·合 0.09	1合·1不 0.42	不·不 0.49
方案 δ_1' 损失值 R	a_1 0	a_1 0	a_1 0	a_1 2 000	a_1 2 000	a_1 2 000
风险值 P	0			2 000		
贝叶斯风险 B	1 000					
方案 δ_2' 损失值 R	a_1 0	a_1 0	a_2 1 500	a_1 2 000	a_1 2 000	a_2 0
风险值 P	60			1 020		
贝叶斯风险 B	540					

续 表

方案 δ_3'	a_1	a_2	a_1	a_1	a_2	a_1
损失值 R	0	1 500	0	2 000	0	2 000
风险值 P		480			1 160	
贝叶斯风险 B			820			
方案 δ_4'	a_2	a_1	a_1	a_2	a_1	a_1
损失值 R	1 500	0	0	0	2 000	2 000
风险值 P		960			1 820	
贝叶斯风险 B			1 390			
方案 δ_5'	a_1	a_2	a_2	a_1	a_2	a_2
损失值 R	0	1 500	1 500	2 000	0	0
风险值 P		540			180	
贝叶斯风险 B			360			
方案 δ_6'	a_2	a_1	a_2	a_2	a_1	a_2
损失值 R	1 500	0	1 500	0	2 000	0
风险值 P		1 020			840	
贝叶斯风险 B			930			
方案 δ_7'	a_2	a_2	a_1	a_2	a_2	a_1
损失值 R	1 500	1 500	0	0	0	2 000
风险值 P		1 440			980	
贝叶斯风险 B			1 210			
方案 δ_8'	a_2	a_2	a_2	a_2	a_2	a_2
损失值 R	1 500	1 500	1 500	0	0	0
风险值 P		1 500			0	
贝叶斯风险 B			750			

因为 $x=$“合・不”与 $x=$“不・合”给决策问题补充的情报信息完全一致，因此采取的行动也应相同。表中两者合记为 $x=$“1 合・1 不”。下面以 δ_4' 为例说明表 13-4 的计算过程。

$$\delta'_4(x)=\begin{cases}a_2 & x=\text{“合·合”}\\ a_1 & x=\text{“不·不”}\\ a_1 & x=\text{“1合·1不”}\end{cases}$$

故

$$P(\theta_1,\delta'_4(x))=\begin{cases}1\,500 & x=\text{“合·合”}\\ 0 & x=\text{“不·不”}\\ 0 & x=\text{“1合·1不”}\end{cases}$$

容易证明:

$$P(\text{合·合}\mid\theta=\theta_1)=0.64,\ P(\text{不·不}\mid\theta=\theta_1)=0.04$$
$$\begin{aligned}P(\text{1合·1不}\mid\theta=\theta_1)&=P(\text{不·合}\mid\theta=\theta_1)+P(\text{合·不}\mid\theta=\theta_1)\\&=0.16+0.16=0.32\end{aligned}$$

故

$$\begin{aligned}R(\theta_1,\delta'_4)&=\sum_{i=1}^{k}R(\theta_1,\delta_4(x_i))P(x=x_i\mid\theta=\theta_1)\\&=150\times0.64+0\times0.04+0\times0.32=960(\text{元})\end{aligned}$$

用同样的方法可求得

$$P(\theta_2,\delta'_4)=1\,820(\text{元})$$

得 δ'_4 的贝叶斯风险为

$$B(\delta_4)=960\times0.5+1\,820\times0.5=1\,390(\text{元})$$

用贝叶斯风险的大小来衡量,决策方案 $\delta'_5(x)$ 为

$$\delta'_5(x)=\begin{cases}a_1 & x=\text{“合·合”}\\ a_2 & x=\text{“不·不”,“1合·1不”}\end{cases}$$

它的贝叶斯风险最小,故它优于其他的 7 个决策方案。

贝叶斯原则是贝叶斯风险最小的决策方案为最佳决策方案。

如上面的例子中,当抽取一件产品检验后而做决策,决策方案为最佳决策方案,若抽取两件产品检验后而做决策,方案 δ'_5 为最佳决策方案。

若我们不做抽样,只凭空猜测,猜设备正常的平均损失为 1 000 元,猜设备不正常的平均损失为 750 元,因此当猜设备不正常,此时平均损失为 750 元。若通过抽取一件产品检验,然后做决策,并采用决策方案 δ_1,这时的平均损失(即贝叶斯风

险)为 450 元。若通过抽取两件产品检验,而后做决策,并采用方案 δ_5',这时的平均损失为 360 元。这些效果完全是由于抽样观察的结果,或者说是补充情报的结果,可称之为补充情报的价值。例如,抽取一件产品检验的补充情报价值为

$$750 - 450 = 300(\text{元})$$

抽取两件产品的补充情报价值为

$$750 - 360 = 390(\text{元})$$

前面我们讲过了完全情报的价值,它与补充情报的关系如下。

【定理 13-1】 任何补充情报的价值都是非负的,且不超过完全情报的价值。

为了简便,我们设情报值(或调查结果)只有 K 种,分别记为 $x_1, x_2, \cdots, x_k$。若我们取得情报 x_i,按决策方案 $\delta(x)$采取行动 $\delta(x_i)$,那么在状态 $\theta = \theta_j$ 时的损失值为$R(\theta_j, \delta(x_i))$,而这时 θ_j 发生的概率为后验概率 $P(\theta_j \mid x = x_i)$,故各种状态下的平均损失为

$$R(\delta(x_i)) = \sum_{j=1}^{n} R(\theta_j, \delta(x_i)) P(\theta_j \mid x = x_i)$$

我们称为后验损失。又利用全概率公式:

$$P(x = x_i) = \sum_{j=1}^{n} P(\theta_j \mid x = x_i) P(\theta_j)$$

则 $P(x = x_i)$ 表示在各种自然状态下情报值为 x_i 的平均概率。

【定理 13-2】 记 $\boldsymbol{B(\delta)}$ 为决策方案 $\boldsymbol{\delta(x)}$ 的贝叶斯风险,则

$$\boldsymbol{B(\delta) = \sum_{i=1}^{k} R(\delta(x_i)) \cdot P(x = x_i)}$$

【定理 13-2】为我们计算贝叶斯风险提供了另外一种方法,用这个计算公式容易求出最佳决策方案。

13.4 最佳决策方案

下面我们按贝叶斯原则求解最佳决策方案。对任何一个决策问题求解最佳决策方案,都有两种基本方法:第一种方法称为正序分析;第二种方法称为反序分析。反序分析所需的计算量少且容易掌握,而且理论上可以证明,反序分析方法求得的最佳决策方案和正序分析求得的最佳决策方案是一致的,故反序分析较常用。下面将主要介绍反序分析方法。

13.4.1 反序分析

反序分析的过程为：在抽样之前，针对所有可能出现的抽样结局(抽样方法与样本容量均事先拟定)，分别计算各自然状态的后验概率，利用这些概率求出各行动方案的后验损失值，然后比较这些后验损失值的大小，选择各种抽样结果下的最佳行动方案，综合成最佳决策方案。下面用一个实例来说明反序分析的全过程。

【例 13－4】 某出口公司的产品每 1 000 件装成一集装箱运交顾客。每箱的不合格品率可分为 5%以下、5%到 15%之间、15%以上 3 种情况。为了计算简便，这 3 种状态分别表示为 $\theta_1 = 0.05$，$\theta_2 = 0.10$，$\theta_3 = 0.15$。按照以往的经验，公司的决策者推测 θ 为这三种值的概率分别为 0.60，0.30，0.10，即先验概率分布为

$$P(\theta = 0.05) = 0.60,\ P(\theta = 0.15) = 0.30,\ P(\theta = 0.15) = 0.10$$

该公司的每箱产品在运交顾客之前，面临这样的决策问题：或是检验箱中每件产品，或是不做任何检验。假如整箱检验，每一件的检验费用为 0.1 元，于是一箱的检验费为 100 元。记这种检验方案为 a_1，采用 a_1 的优点是可检查出一箱中的所有不合格品，保证运交顾客的产品百分之百合格。假如整箱都不做检验，顾客买到不合格品时必须准予更换，每更换的一件所需的费用总和为 1.25 元。这种整箱都不检验的行动方案记为 a_2，采用 a_2 的优点是可节省检验费 100 元。

决策者从每箱中抽出两件产品检验，并通过这两件产品提供的情况做出最佳决策方案，方法如下：

设抽出的两件产品为 z_1，z_2，并规定当第 i 件产品检验结果为不合格品时，记 $z_i = 1$，否则记 $z_i = 0$，另外设抽样总的结果为

$$x = z_1 + z_2$$

x 的值刚好为抽出的两件产品中的不合格品数目，它是一个随机变量。在抽样试验前，我们就知道 x 可能有 0，1，2 三种结果。x 的概率分布是超几何分布，此处可认为 x 近似服从二项分布。

在实际进行抽样试验前，决策者先做下面分析：假定抽样后观察到的不合格品数为 0，即 $x = 0$，则可计算各状态 θ_1，θ_2，θ_3 的后验概率及每一种可能行动方案 a_1，a_2 的后验损失，计算格式及结果如表 13－5 所示。表中数值显示行动 a_1 的后验损失为 100 元，而行动 a_2 的后验损失为 90.75 元，故出现 $x = 0$ 的抽样结果时，最佳行动方案为 a_2。假设抽样结果为 $x = 1$，计算结果列入表 13－6 中可得最佳行动方案为 a_1；最后假设抽样结果为 $x = 2$，计算结果列入表 13－7 中可得最佳行动方案为 a_1。综合这些结论，即得最佳决策方案为

$$\delta(x)=\begin{cases}a_1 & x=1,2\\ a_2 & x=0\end{cases}$$

表 13-5 不合格品为 0 件(即 $x=0$) 时后验损失表

自然状态 θ	先验概率 $P(\theta)$	各状态下 $x=0$ 的概率 $P(x=0\mid\theta)$	后验概率 $P(\theta\mid x=0)$	各行动方案的损失 $R(\theta, a)$	
				a_1	a_2
0.05	0.6	0.902	0.632	100	62.50
0.10	0.3	0.810	0.284	100	125.00
0.15	0.1	0.722	0.084	100	187.50
后验损失 $\sum_{i=1}^{3}R(\theta_i, a)\cdot P(\theta_i\mid x=0)$				100	90.75

表 13-6 不合格品为 1 件(即 $x=1$) 时后验损失表

自然状态 θ	先验概率 $P(\theta)$	各状态下 $x=1$ 的概率 $P(x=1\mid\theta)$	后验概率 $P(\theta\mid x=1)$	各行动方案的损失 $R(\theta, a)$	
				a_1	a_2
0.05	0.6	0.095	0.418	100	62.5
0.10	0.3	0.180	0.396	100	125.0
0.15	0.1	0.255	0.186	100	132.62
后验损失 $\sum_{i=1}^{3}R(\theta_i, a)\cdot P(\theta_i\mid x=1)$				100	110.5

表 13-7 不合格为 2 件(即 $x=2$) 时后验损失表

自然状态 θ	先验概率 $P(\theta)$	各状态下 $x=1$ 的概率 $P(x=2\mid\theta)$	后验概率 $P(\theta\mid x=2)$	各行动方案的损失 $R(\theta, a)$	
				a_1	a_2
0.05	0.6	0.002	0.194	100	62.5
0.10	0.3	0.010	0.490	100	125.0
0.15	0.1	0.022	0.316	100	187.5
后验损失 $\sum_{i=1}^{3}R(\theta_i, a)\cdot P(\theta_i\mid x=2)$				100	132.62

下面我们将简单证明，由反序分析求得的最佳方案确为贝叶斯原则下的最佳决策方案。

设某决策问题的自然状态为 θ_1，θ_2，…，θ_m；先验概率分别为 $P(\theta_1)$，$P(\theta_2)$，…，

$P(\theta_m)$；可采取行动为 a_1，a_2，…，a_n；而当状态为 θ_j 时采取行动 a_i 的损失为 $R(\theta_i, a_j)$，我们便补充信息，例如进行抽样，设补充信息 x 的所有可能的结果为 x_1，x_2，…，x_k，记 x_i 在状态 θ_j 下发生的概率为 $P=(x=x_i \mid \theta_j)$，又记：

$$P(x=x_i)=\sum_{j=1}^{m} P=(x=x_i \mid \theta_j)P(\theta_j) \quad (i=1, 2, \cdots, k)$$

若 $\delta(x)$ 为任一决策方案，而 $\delta_0(x)$ 是按反序分析得到的最佳决策方案，当 $x=x_i$，按照这两个决策方案，我们分别应采取行动为 $\delta(x_i)$ 和 $\delta_0(x_i)$，由反序分析过程可知，后验损失为

$$R(\delta_0(x_i)) \leqslant R(\delta_1(x_i)) \quad (i=1, 2, \cdots, k)$$

故贝叶斯风险为

$$\begin{aligned} B(\delta_0) &= \sum_{i=1}^{k} R(\delta_0(x_i)) \cdot P(x=x_i) \\ &\leqslant \sum_{i=1}^{k} R(\delta_1(x_i)) \cdot P(x=x_i) \\ &= B(\delta) \end{aligned}$$

由于 $\delta(x)$ 是任意选取的一个决策方案，上式即说明在所有的决策方案中，方案 $\delta_0(x)$ 贝叶斯风险小。

下面用决策树来表示反序决策过程，仍以【例 13－4】为例，得到如图 13－1 所示的决策树。

在决策树中，在各种状态变量下，所抽取的两件产品中没有一件不合格品的概率 P 可由全概率公式计算如下：

$$\begin{aligned} P &= P(x=0) \\ &= P(x=0 \mid \theta=0.5)P(\theta=0.5)+P(x=0 \mid \theta=0.10)P(\theta=0.10)+ \\ &\quad P(x=0 \mid \theta=0.15)P(\theta=0.15) \end{aligned}$$

用二项分布近似计算：

$$P(x=0 \mid \theta=0.5)=(0.95)^2=0.9025$$
$$P(x=0 \mid \theta=0.10)=(0.90)^2=0.81$$
$$P(x=0 \mid \theta=0.15)=(0.85)^2=0.7225$$

故得 $P=P(x=0)=0.9025 \times 0.6+0.81 \times 0.3+0.7225 \times 0.1=0.857$。

仿此可计算 $P(x=1)$ 及 $P(x=2)$。

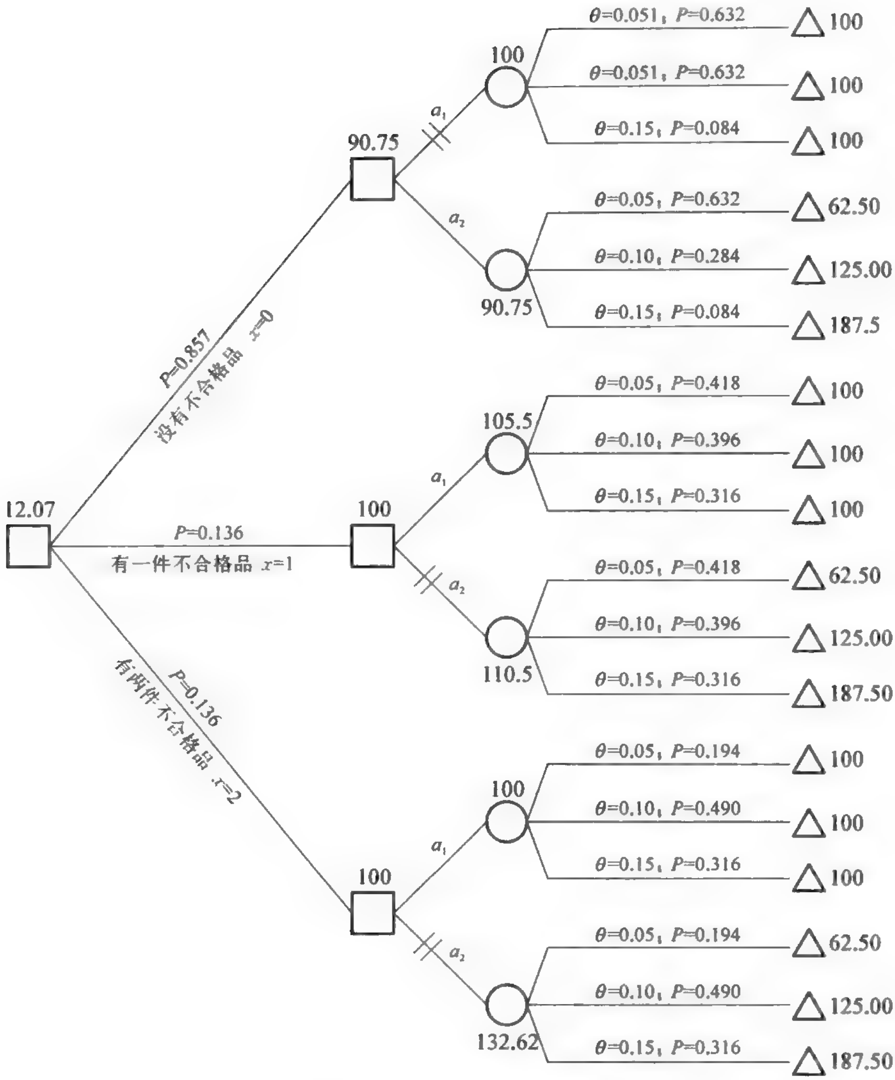


图 13-1 反序分析决策树

13.4.2 正序分析

正序分析的过程为：在进行抽样或其他调查之前，枚举所有可能出现的抽样结果 x 及所有可能的决策方案 $\delta(x)$，针对每一种方案 $\delta(x)$ 计算其贝叶斯风险，比较各种方案的贝叶斯风险的大小，贝叶斯风险最小的决策方案即为最佳决策方案。

例如，13.3 节中的例子即是采用正序分析。

13.5 最佳样本容量

我们已经知道,抽样调查可以获得补充情报,以减少不确定性代价,改善决策效果。但是与获得其他类型的补充情报一样,一般来说,是要支付一定费用的。所以,对于一个具体的决策问题,是否要抽样?如果抽样,样本容量又应当是多大?这是抽样之前必须弄清楚的问题。

抽样所支付的费用称为抽样成本。样本容量为 N 时的抽样成本记为 $C(N)$,显然有:

$$C(0)=0$$

若 $N\neq 0$,抽样成本也可以分为两部分:固定成本和可变成本。用 C_{j} 表示固定成本;一般情况下,可变成本与 N 成正比,用 C_{v} 表示单位可变成本,则有:

$$C(N)=C_{\mathrm{j}}+C_{\mathrm{v}}N\ (N\neq 0)$$

当样本容量 N 确定以后,抽样情报价值也随之确定。抽样情报价值也是 N 的函数,记之为 $EVSI(N)$。对不同的 N,抽样情报价值可以不同。

称下面的差数:

$$ENGS(N)=EVSI(N)-C(N)$$

为抽样净值。它反映出在扣除抽样成本之后,抽样给决策带来的纯净的好处。

显然,如果对任何自然数 N 都有 $ENGS(N)\leqslant 0$,则不宜进行抽样;如果对某一个 N 有 $ENGS(N)>0$,则进行容量为 N 的抽样是值得的(当然,不一定是最好的)。使 $ENGS(N)$ 达到最大值的非负整数称为最佳样本容量。

由于求解最佳样本容量的计算量相当大,人工计算相当困难,要借助电子计算机才能奏效。这里只简单说明计算方法:取 $N=1$ 计算 $ENGS(1)<0$,则不宜抽样;若 $ENGS(1)>0$,则取 $N=2$,计算 $ENGS(2)$,$ENGS(2)<0$,则最佳样本容量为 1;若 $ENGS(2)>0$,取 $N=3$,计算 $ENGS(3)$,若 $ENGS(3)<0$,则停止计算,找出 $ENGS(1)$、$ENGS(2)$ 中最大者,其对应的 N 为最佳样本容量;若 $ENGS(3)>0$,继续 $N=4$,……,如此下去。已有人证明,一定存在 K,使 $ENGS(K)<0$。比较 $ENGS(1)$,$ENGS(2)$,…,$ENGS(K-1)$,其中最大者对应的 N 即为最佳样本容量。

思考与练习

1. 某厂的甲、乙两车间生产同一种产品,产量分别是每天 300 件和 500 件。已

知甲车间产品合格率为90%，乙车间合格率为95%。求此工厂这种产品的合格率。

2. 某工厂有一套自动生产设备，每次生产前，该设备需要经过精密调整，以确保质量。依以往经验可知：若该设备调整良好，则生产的产品有95%为合格品；若不成功，则仅有30%的产品为合格品。此外由以往的记录，调整成功的次数占75%。该厂做完一次调整工作后，就试生产数件，以判断设备有没有调整好。若试生产一件产品，发现其为不合格品，试据此判断设备是否已调好。

3. 某公司准备经营某种新产品，可以采取的行动有：小批生产试销、中批生产及大批生产。可能出现的销售状态有：畅销、一般及滞销。如大批生产，在畅销时可获利100万元，一般时可获利30万元，滞销时亏损60万元；如中批生产，在前两种情况下分别获利50万元、40万元，滞销时则亏损20万元；如小批生产，则在3种情况下分别获利10万元、9万元、6万元。又根据长期经验，同类产品为畅销、一般及滞销的概率分别是0.2，0.5，0.3，这对于现在正在考虑的新产品来说，可以作为先验分布。从某些迹象看来，这种新产品的市场需求情况似有变化，为了弄清这点，进行了市场预测。预测的准确程度如题表13-1所示。

题表 13-1

θ(状态)	$P(\theta)$	$P(H_1\|\theta)$	$P(H_2\|\theta)$	$P(H_3\|\theta)$
θ_1(畅销)	0.2	0.80	0.15	0.05
θ_2(一般)	0.5	0.20	0.70	0.10
θ_3(滞销)	0.3	0.02	0.08	0.90

其中H_1、H_2、H_3分别表示预测结果为畅销、一般、滞销。

(1) 如果预测结果为畅销，试进行后验分析，并根据后验分布求出该公司的最优方案。

(2) 如果预测结果为一般或滞销，最优行动又是什么？写出最优决策方案。

4. 某食品公司考虑是否把一种新产品投放市场。公司决策者估计，这种新产品受消费者欢迎的概率为0.6，不受欢迎的概率为0.4。如果受欢迎，则可获利40 000元；如果不受欢迎，则亏损35 000元。如果在决定是否产销这种新产品之前进行市场调查，调查结果有受欢迎和不受欢迎两种。调查结果与实际情况之间关系如题表13-2所示。

题表 13-2

实际情况	调查结果为受欢迎的概率	调查结果为不受欢迎的概率
受欢迎	0.8	0.2
不受欢迎	0.1	0.9

市场调查费用为 10 000 元。那么，该公司若根据经验的知识(即先验概率)，不进行市场调查，应如何决定？若进行市场调查，又应如何决策？是否有必要做市场调查？

14 多目标决策法

14.1 多目标决策概述

14.1.1 多目标决策的特点

两个目标以上的决策称为多目标决策。在实际生产经营活动中，多目标问题是经常遇到的。例如，对于港口企业生产管理，不仅要追求一个理想的收入水平，同时还要设法降低服务成本、提高服务效率、减少对环境的污染等。有些决策问题本身明显地要求多目标，如一个港口城市的整体经济发展就是一个复杂的系统，而这个系统的活动就是一个多目标的活动过程，其子系统和各地方、各部门、各企业的经济活动都是多目标活动。因此当研究系统时，例如研究港口经济综合效益，就具有多目标的特点。有些决策问题，往往开始似乎只考虑一个目标，可是在进一步考虑具体措施时，还得考虑这个措施执行之后可能会出现哪些问题，并对其提出一定的要求或准则，这样就等于把防止这些出现的问题也列为目标，从而使单一目标决策变为多目标决策问题了。

多目标决策问题的共同特点如下：

1）目标之间的不可公度性

目标之间的不可公度性，是指各个目标没有统一的衡量标准，因而很难进行比较。例如，外贸出口商品中，数量有各种计量，如台、吨、件等，商品质量又有合格率等。若在多目标决策问题中，能找到某个方案，使所有目标都能达到最优，那么不可公度性也就不成问题了。但这种情况很少发生。

2）目标之间的矛盾性

大部分目标决策问题存在着矛盾，即如果采用这一种方案去改进某一目标的值，很可能会使另一目标的值变坏。如建码头，若提高码头装卸效率，就要增加码头泊位、增加装卸机械数量，但是这就会需要增加大量投资。

14.1.2 多目标决策的基本组成要素

1）决策单元

在决策过程中，决策人、决策分析人员和电子计算机等结合起来构成决策单

元。其主要作用是：收集并处理各种信息，使之成为系统的知识；制定决策规则；做出决定等。

2）目标集

人们都希望达到一定的目的，这个目的反映着决策者的要求和愿望，这就是目标。目标有多层次：小目标、中目标、总目标等。最高层的目标是促使人们研究这个决策问题的原动力，但是这个目标常常太笼统，不便于计算。小目标比中目标更便于运算。

3）属性

所谓属性是用来表示目标达到程度的评价标准。对属性的要求是能够测量并且能够易于理解。例如，研究目标是"充分利用码头资源"，就可用"机械每小时作业量"这一属性来表示。

14.1.3 多目标决策的基本原则

由于目标的增多，使得所要分析的问题变得复杂化，因此在处理多目标决策问题时，应当根据决策需要，尽量减少决策目标数。一般要遵循两个基本原则。

14.1.3.1 化多为少原则

因为目标数越多，选择标准就越多，比较和选择各种不同方案就越困难。因此，在实际问题中，应将目标化多为少。即在满足决策的前提下，尽量减少目标的个数。通常的做法有如下几种：

1）剔除办法

剔除那些不必要和从属性的目标。例如，已把"提高利润"作为决策的目标，就不必再将"降低成本"也同时列为决策目标。因为降低成本就是实现提高利润的手段之一，是从属性的子目标，可以剔除。如果决策的各目标中，包括两个对立而无法协调的目标，经过决策者权衡之后，在必要时，就应牺牲其中的一个。

2）合并办法

多目标决策问题由于目标之间有明显的客观联系，故可以把类似的几个目标合并为一个目标来解决。如一个物流运输企业要求做到降低运输成本费用、降低管理费用、降低人工费用等，就可以把它们合并成为"降低成本"一个目标。

3）把次要目标列为约束条件办法

根据各个目标的重要性，分清主次关系，把本质的主要目标列为目标，而把其余的非主要、非本质的列为约束条件。

4）构成综合目标办法

把几个目标通过同度量、平均或构成函数的办法构成一个综合目标。用函数表示为

$$P = f(P_1, P_2, \cdots, P_n)$$

式中，P 表示综合目标；$P_1, P_2, \cdots, P_n$ 表示子目标。

14.1.3.2 目标排序法

目标排序法就是决策者根据目标的重要程度排成一个次序，最重要的目标排在第一位，在选择方案时，必须先达到重要目标后才能再考虑下一个目标，然后再进行选择，做出决策。

在实际工作中，遇到多目标决策问题时，虽然可以通过一些办法把目标数目减少，化为单一目标问题来解决。但是在有些情况下却很难做到这一点，难以用简单的方法按各目标间的客观联系直接转化为单目标，因此仍然需要寻找解决多目标决策问题的方法。

14.2 目标规划法

14.2.1 基本概念

目标规划也称多目标规划，是运筹学中应用面较广的一个重要分支。早在 20 世纪 60 年代由美国数学家查纳(Charnes)和库伯(Cooper)等人提出，以后不断完善；70 年代引起人们的普遍重视，在很多领域有广泛的应用，并取得了显著的经济效果。

目标规划法是在线性规划的基础上发展起来的，既保持了线性规划易于计算的特点，也克服了线性规划只能解决单一目标优化问题的局限性。它是解决与协调各种约束条件和目标之间重要程度不同的多目标决策问题的一个有效工具。它的基本含义是，求一组非负变量，在满足一定线性约束(资源约束)与多个线性目标约束条件下，实现计划管理目标与实际可能完成的目标之间的偏差总和为最小。

14.2.2 目标规划的数学模型

14.2.2.1 确定决策变量和设计偏差变量

目标规划模型与线性规划一样，除了要确定决策变量(非负)之外，重要的是要引入反映偏差的偏差变量，用符号 d^-，d^+ 表示，且 d^-，d^+ 为非负数，d^+ 表示超出完成目标或可供资源的偏差变量，d^- 表示低于完成目标或可供资源的偏差变量，在同一目标或资源限制的偏差变量 d^+ 与 d^- 中，至少有一个为 0，即有 $d^+ \cdot d^- = 0$。

14.2.2.2 目标函数和目标等级

由于目标规划是解决多目标决策问题，它并不能像线性规划那样直接求出单

目标决策的最优解，而是要求在给定的约束集合内，使计划目标和实际可能达到的目标值之间的偏差总和为最小。所以，它的目标函数是通过给定的目标的偏差变量来描述的，要求其最小值。

目标规划法在解决多目标决策时，关键的一步是要将目标按其重要程度划分为不同等级，即目标的优先级。优先级是与目标的重要性相联系的，同等重要的目标划归为一类，放在同一优先级内。重要性越大的目标，其优先级越高。另外，同一优先次序的偏差变量的度量单位必须相同，不同优先次序的偏差变量的度量单位可以不同。

为了把优先级概念引入模型中，需要设计优先级因子，一般用符号 P 表示，并且用 P_i 表示第 i 级优先因子，同时约定：

$$P_1 \gg P_2 \gg \cdots \gg P_i \gg P_{i+1}$$

符号“$\gg$”表示“远远大于”的意思，即对于任意一个多么大的正数 K 与 P_{i+1} 相乘，都不可能得到 $KP_{i+1} \geqslant P_i$ 的结果（$i = 1, 2, \cdots$）。模型中的优先次序一旦被确定下来，在求解过程中，一定要严格做到使最重要目标的未达到部分尽可能最小，然后再考虑次一级目标。

目标规划的目标函数是通过给定目标的偏差变量来描述的，但是偏差变量的引入是否正确，一般要遵守以下几个原则：

（1）目标要求恰好达到，则将该项目标的正负偏差均列入目标函数，此时目标函数为

$$\text{极小化} \quad Z = d_i^+ + d_i^-$$

（2）目标要求超额完成，如总产值、利润等，则应将目标的负偏差列入目标函数，此时目标函数为

$$\text{极小化} \quad Z = d_i^-$$

（3）目标要求不要超过，如机器设备运转时间、车船在港等待时间等，则应将目标的正偏差列入目标函数，此时目标函数为

$$\text{极小化} \quad Z = d_i^+$$

（4）对于目标函数同一级别的偏差变量，由于决策者的要求不同，可以给偏差变量赋给不同的权数，来区分其重要程度。但必须做到偏差变量所表示的目标度量单位一致，这称为多目标中的分目标结构。

所以，目标规划的目标函数的一般模式为

$$\text{极小化} \quad Z = \sum_{i=1}^{g} P_i(w_1 d_i^- + w_2 d_i^+)$$

式中，w_1、w_2 为同一目标级别的偏差变量的权数。

14.2.2.3 目标约束条件和资源约束条件

目标规划的约束条件分为目标约束条件和资源约束条件，一般形式为

$$\sum_{j=1}^{n} a_{ij} x_j + d_i^- - d_i^+ = g_i$$

式中，x_j 表示决策变量；g_i 表示给定的目标水平（$i = 1, 2, \cdots, g$）；a_{ij} 表示决策变量与目标水平 g_k 有关的技术系数；d_i^+、d_i^- 表示以目标水平 g_i 为标准而引入的偏差变量。

一般约束（资源条件）和线性规划一样，根据要求列出不等式组，但若已列入目标约束就不要重复，以目标约束为主。

综上所述，目标规划的数学模型为

$$\text{极小化} \quad \text{目标函数}\ Z = \sum_{i=1}^{g} P_i(w_1 d_i^- + w_2 d_i^+)$$

$$\text{约束条件：} \sum_{j=1}^{n} a_{ij} x_j + d_i^- - d_i^+ = g_i$$

$$\text{一般约束：} \sum_{j=1}^{n} b_{ij} x_j (\leqslant, =, \geqslant) c_j$$

$$\text{非负条件：} x_j, d_i^-, d_i^+ \geqslant 0$$

14.2.3 目标规划应用实例

【例 14－1】 某码头有 A、B 两种货物装船，需要岸边和堆场两种港口机械，每班的生产能力以及两种货物生产工时定额和单位收入如表 14－1 所示。

表 14－1 A、B 两种货物装船情况

工序 定额	工时定额		生产能力/(小时/班)
	货物 A	货物 B	
岸边机械	1	3	60
堆场机械	1	1	40
产值(元/件)	15	25	

现管理部门提出三级目标：

一级目标是每班产值 750 元，尽量超值完成；

二级目标是充分利用两种港口机械（已知岸边机械工时费用为堆场机械工时费用的 2 倍）；

三级目标是尽量减少两种港口机械的加班工时。

现要求应用目标规划方法在争取上述管理目标逐级实现的条件下，拟订一个满意的生产规划。

【解】 1）确定决策变量和设计偏差变量

设 x_1、x_2 分别为货物 A、B 的装船量，d_i^+、d_i^- 为不同目标的偏差变量，且 x_1，x_2，d_i^+，$d_i^- \geqslant 0$。

2）约束条件

（1）目标约束：

$$P_1 \text{ 级：} 15x_1 + 25x_2 + d_1^- - d_1^+ = 750$$

d_1^+、d_1^- 分别为每班产值 750 元的目标而设立的正负偏差变量，由于目标要求超额完成，所以应将 d_1^- 列入目标函数，即 $P_1 d_1^-$。

$$P_2 \text{ 级：岸边机械工时，} x_1 + 3x_2 + d_2^- - d_2^+ = 60$$

$$\text{堆场机械工时，} x_1 + x_2 + d_3^- - d_3^+ = 40$$

d_2^-、d_2^+ 和 d_3^-、d_3^+ 分别是两种机械的工时为目标的偏差变量，两种机械的工时都争取充分利用，故应将 d_2^-、d_3^- 分别列入目标函数，但考虑码头生产成本的问题，两个偏差变量按费用的比例（1∶2）为权数，开辟一个分目标结构，即 $P_2(d_2^- + 2d_3^-)$。

P_3 级：在目标约束中对 d_2^+、d_3^+ 表示超出充分利用两种机械工时后的加班工时，故只需将对 d_2^+、d_3^+ 列入目标函数即可，与 P_2 级同理，开辟一个分目标结构，即 $P_3(d_2^+ + 2d_3^+)$。

（2）本例一般条件约束：无。

3）目标函数

$$\text{极小化 目标函数 } Z = P_1 d_1^- + P_2(d_2^- + 2d_3^-) + P_3(d_2^+ + 2d_3^+)$$

综合本例，目标规划的数学模型为

$$\text{极小化 目标函数 } Z = P_1 d_1^- + P_2(d_2^- + 2d_3^-) + P_3(d_2^+ + 2d_3^+)$$

$$\text{约束条件}\begin{cases}15x_1+25x_2+d_1^- - d_1^+ = 750\\ x_1+3x_2+d_2^- - d_2^+ = 60\\ x_1+x_2+d_3^- - d_3^+ = 40\end{cases}$$

非负条件：$x_1, x_2, d_i^-, d_i^+ \geqslant 0$

14.2.4 目标规划的求解方法

目标规划的解法有两种，一种是单纯形方法，用于解多级目标和任意变量，但其过程较复杂，这里不做介绍。另一种是图解法，这种解法比较直观，但只适用于求解两个决策变量的多目标规划问题，现结合实例简要介绍其求解过程。

【例 14-2】 某装配企业有甲、乙两种产品装配，每种产品的单位耗用人力和设备台时及生产价值如表 14-2 所示。

表 14-2 甲、乙两种产品装配情况

资 源	产 品		
	甲	乙	单 位
单位耗用人力时间	6	8	小时/件
单位耗用设备台时	10	6	台时/件
单位价值	80	100	元/百米

第一级要求人力工作不超过 480 小时，第二级要求设备工作不超过 580 台时，第三级要求产值不少于 8 000 元。求出目标规划的解。

【解】 设 x_1、x_2 为甲、乙两产品的产量，d_i^+、d_i^- 为不同级别目标的偏差变量，且 x_1、x_2，d_i^+、$d_i^- \geqslant 0\ (i=1, 2, 3)$。

$$P_1\text{ 级：}6x_1+8x_2+d_1^- - d_1^+ = 480$$

d_1^-、d_1^+ 分别为人力工作不超过 480 小时目标而设立的偏差变量，由于目标要求不超过，所以应将 d_1^+ 列入目标函数，即 $P_1 d_1^+$。

$$P_2\text{ 级：}10x_1+6x_2+d_2^- - d_2^+ = 580$$

d_2^+、d_2^- 分别为设备工作不超过 580 小时目标而设立的偏差变量，但由于设备不能停止运转，以免闲置，所以目标要求恰好达到，因此应将 $d_2^- + d_2^+$ 列入目标函数，即 $P_2(d_2^- + d_2^+)$。

$$P_3\text{ 级：}80x_1+100x_2+d_3^- - d_3^+ = 8\,000$$

d_3^+、d_3^- 分别为产值不少于 8 000 元而设立的偏差变量，由于目标要求不少于 8 000元，即希望产值超额完成. 所以应将 d_3^- 列入目标函数，即 $P_3d_3^-$。

故本题的目标函数为

$$\text{极小化}\quad \text{目标函数 } Z = P_1d_1^+ + P_2(d_2^- + d_2^+) + P_3d_3^-$$

$$\text{约束条件}\begin{cases}6x_1 + 8x_2 + d_1^- - d_1^+ = 480\\ 10x_1 + 6x_2 + d_2^- - d_2^+ = 580\\ 80x_1 + 100x_2 + d_3^- - d_3^+ = 8\,000\end{cases}$$

$$\text{非负条件：} x_1、x_2,\ d_i^-、d_i^+ \geqslant 0\ (i = 1,\ 2,\ 3)$$

采用图解法求解：

建立直角坐标系，如图 14－1 所示，设横坐标为 x_1，纵坐标为 x_2。因本题无一般约束条件，故可行域为第Ⅰ象限。

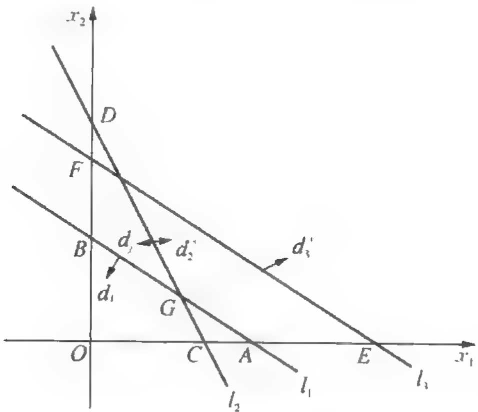


图 14－1 图解法示意图

若不考虑偏差变量，则根据约束条件可得三个方程：

$$\begin{cases}6x_1 + 8x_2 = 480 & (l_1)\\ 10x_1 + 6x_2 = 580 & (l_2)\\ 8x_1 + 100x_2 = 8\,000 & (l_3)\end{cases}$$

作出 3 个方程的直线如图 14－1 中的 l_1、l_2、l_3 所示，则 P_1 级目标希望不超额计划用量，d_1^+ 要趋于最小，所以 P_1 级满意域为三角形 OAB 内部及其边界。

P_2 级目标希望达到，则 P_2 级的满意域为线段CD。

P_3 级目标希望能超额完成，d_3^- 趋于最小，所以 P_3 级的满意域为线段 EF 外

以 d_3^+ 指明的一侧。

对上面的满意域逐个地进行叠加，P_1 级和 P_2 级叠加的结果满意域为线段 CG，但 CG 与 P_3 级满意域不相交。所以应从 P_1、P_2 级满意域——线段 CG 中找出离 P_3 级满意域最近的点，即为 G 点，所以该目标规划的解是直线 l_1 与 l_2 的交点 G，即

$$\begin{cases}6x_1 + 8x_2 = 480 \\ 10x_1 + 6x_2 = 580\end{cases}$$

解得 $G(40, 30)$。

将 $x_1 = 40$，$x_2 = 30$，代入 P_1 级目标约束：

$$6x_1 + 8x_2 + d_1^- - d_1^+ = 480$$

得 $d_1^+ = 0$，即说明人力工作没有超过，P_1 级目标达到。

代入 P_2 级目标约束 $10x_1 + 6x_2 + d_2^- - d_2^+ = 580$，有 $d_2^- = d_2^+ = 0$，即恰好利用设备能力，P_2 级目标达到。

代入 P_3 级目标约束 $80x_1 + 100x_2 + d_3^- - d_3^+ = 8\,000$，得 $d_3^- = 1\,800$，说明产值还缺 1 800。但从坐标图可知，在满足 P_1 级和 P_2 级目标的前提下，$d_3^- = 1\,800$ 是极小值。所以通常称 G 为满意解。

14.3 层次分析法(AHP)

14.3.1 AHP 概述

层次分析法是一种实用的多准则决策方法。它把一个复杂问题表示为一个有序的递阶层次结构，利用人们的判断，对决策方案的优劣进行排序。这种方法能够统一处理决策中的定性与定量因素，具有实用性、系统性、简洁性等优点。

层次分析法是美国著名运筹学家、匹兹堡大学教授 T. L. Saaty 于 20 世纪 70 年代中期提出的。Saaty 在 1971 年曾为美国国防部研究“应急计划”，1972 年为美国科学基金会研究电力在工业部门的分配问题，1973 年为苏丹政府研究苏丹的运输问题。在此期间，由于研究工作的需要，他感到需要有一种方法，可以综合定性与定量分析，使人脑的决策思维过程模型化。1977 年，Saaty 正式提出了 AHP 法。1980 年以来，Saaty 出版了多部专著，全面论述 AHP 的原理、应用及数学基础。

AHP 的基本思路是决策者首先将复杂问题分解为若干组成要素，并将这些要素按支配关系形成有序的递阶层次结构；然后通过两两比较，确定层次中诸要素的

相对重要性;最后综合各层次要素的重要程度,得到诸要素的综合评价值,并据此进行决策。层次分析法体现了人们在决策思维过程中进行分解、判断、综合的基本特征。

14.3.2 AHP的引出

14.3.2.1 一个例子

下面用一个例子来说明 AHP 的基本思路。

假定现在有 n 个西瓜,它们的重量可用一个向量$[W_1, W_2, \cdots, W_n]^{\mathrm{T}}$ 表示。这些西瓜的重量未知,如果要想知道这些西瓜按重量大小的排序情况,应该怎样获得?也就是说,如何才能估计出这些西瓜的相对重量?或者说,如何才能得到它们的重量向量$[W_1, W_2, \cdots, W_n]^{\mathrm{T}}$?

方法一:对于这个问题,很容易想到的一个解决方法是:用一杆秤依次称出各个西瓜的重量 $W_1, W_2, \cdots, W_n$。这样就可对这些西瓜的重量进行比较了。

但是,如果现在没有秤,如何才能估计出西瓜的轻重呢?这时可采用第二种方法。

方法二:采用两两比较的方法,判断每两个西瓜的相对重量的比例。具体做法是:先取出第一个西瓜,将它依次与第二个、第三个、……、第 n 个西瓜进行两两比较(可用两只手掂一掂来比较);然后将第一个西瓜放回原处,取出第二个西瓜,将它与其他西瓜进行两两比较;依此类推,直到将第 n 个西瓜与其他西瓜都进行了两两比较为止,这样就得到了一个反映西瓜两两比较的相对重量的 $n\times n$ 矩阵,称为比较判断矩阵。假设用 $\boldsymbol{A}$ 表示该比较判断矩阵,则

$$\boldsymbol{A}=\begin{bmatrix} \frac{W_1}{W_1} & \frac{W_1}{W_2} & \cdots & \frac{W_1}{W_j} & \cdots & \frac{W_1}{W_n} \\ \frac{W_2}{W_1} & \frac{W_2}{W_2} & \cdots & \frac{W_2}{W_j} & \cdots & \frac{W_2}{W_n} \\ \vdots & \vdots & & \vdots & & \vdots \\ \frac{W_i}{W_1} & \frac{W_i}{W_2} & \cdots & \frac{W_i}{W_j} & \cdots & \frac{W_i}{W_n} \\ \vdots & \vdots & & \vdots & & \vdots \\ \frac{W_n}{W_1} & \frac{W_n}{W_2} & \cdots & \frac{W_n}{W_j} & \cdots & \frac{W_n}{W_n} \end{bmatrix}_{n\times n}$$

$$=(a_{ij})_{n\times n}\quad (i,\ j=1,\ 2,\ \cdots,\ n) \tag{14-1}$$

式(14-1)中，
$$a_{ij}=\frac{W_i}{W_j}\quad(i,\ j=1,\ 2,\ \cdots,\ n)\tag{14-2}$$

显然有：

$$a_{ij}=\frac{1}{a_{ji}}\quad(i,\ j=1,\ 2,\ \cdots,\ n)\tag{14-3}$$
$$a_{ii}=1\quad(i,\ j=1,\ 2,\ \cdots,\ n)\tag{14-4}$$
$$a_{ij}=\frac{a_{ik}}{a_{jk}}\quad(i,\ j,\ k=1,\ 2,\ \cdots,\ n)\tag{14-5}$$

用重量向量 $\boldsymbol{W}=[W_1,\ W_2,\ \cdots,\ W_n]^{\mathrm{T}}$ 右乘等式(14-1)两边，得

$$\begin{aligned}\boldsymbol{AW}&=\begin{bmatrix}\frac{W_1}{W_1} & \frac{W_1}{W_2} & \cdots & \frac{W_1}{W_j} & \cdots & \frac{W_1}{W_n}\\ \frac{W_2}{W_1} & \frac{W_2}{W_2} & \cdots & \frac{W_2}{W_j} & \cdots & \frac{W_2}{W_n}\\ \vdots & \vdots & & \vdots & & \vdots\\ \frac{W_i}{W_1} & \frac{W_i}{W_2} & \cdots & \frac{W_i}{W_j} & \cdots & \frac{W_i}{W_n}\\ \vdots & \vdots & & \vdots & & \vdots\\ \frac{W_n}{W_1} & \frac{W_n}{W_2} & \cdots & \frac{W_n}{W_j} & \cdots & \frac{W_n}{W_n}\end{bmatrix}_{n\times n}\begin{bmatrix}W_1\\ W_2\\ \vdots\\ W_j\\ \vdots\\ W_n\end{bmatrix}\\ &=n\begin{bmatrix}W_1\\ W_2\\ \vdots\\ W_j\\ \vdots\\ W_n\end{bmatrix}\\ &=n\boldsymbol{W}\end{aligned}\tag{14-6}$$

从式(14-6)可见：

(1) 重量向量 $\boldsymbol{W}$ 是比较判断矩阵 $\boldsymbol{A}$ 的特征向量。

(2) 重量向量的元素个数 n(即西瓜的个数)是比较判断矩阵 $\boldsymbol{A}$ 的特征值。

(3) 只要求出判断矩阵 $\boldsymbol{A}$ 的特征向量，则该特征向量就是重量向量 $\boldsymbol{W}$。

从以上分析可知，确定 n 个西瓜的相对重量排序(即重量向量 $\boldsymbol{W}$)有两种方法。第一种方法是用秤逐个称出西瓜的重量，加以比较；第二种方法是用两两比较的方

法构造出比较判断矩阵，然后采用数学方法计算出该比较判断矩阵的特征值与特征向量，该特征向量就是所要求出的重量向量 $\boldsymbol{W}$。

第一种方法简单、直观，但仅适用于有度量标尺的问题。在前面所述的西瓜例子中当然可采用第一种方法，但是在社会经济系统中，很多测度对象具有相对的性质，无法确定统一的标度，例如对环境、安全、舒适这类问题就很难提出一种标度。这时第二种方法就十分有利。因此，第二种方法适用于社会、政治、人的行为、管理等难以度量的问题。在采用第二种方法时，重量向量的含义就演变为重要性向量，或称相对重要性排序权值向量。层次分析法就是采用第二种方法来进行评价与决策的。

14.3.2.2 判断尺度

如上所述，第二种方法的关键之一是构造比较判断矩阵。比较判断矩阵是描述对于上一层次某要素来说本层次相关要素之间相对重要性的矩阵，它是以上一层次某要素为评价准则、对本层次的要素进行两两比较得出的。为了将两两比较的结果数量化，需要有一个表示两个要素的相对重要性的数量尺度，称为判断尺度。层次分析法采用 1～9 标度的判断尺度，其定义如表 14－3 所示。

表 14－3 判断尺度定义表

判断尺度	定 义
1	表示两个要素相比，具有同样的重要性
3	表示两个要素相比，一个要素比另一个要素稍微重要
5	表示两个要素相比，一个要素比另一个要素明显重要
7	表示两个要素相比，一个要素比另一个要素强烈重要
9	表示两个要素相比，一个要素比另一个要素极端重要
2,4,6,8	介于上述两个相邻判断尺度的中间

表中各数的倒数表示否定的意思，例如，如果要素 i 比要素 j 明显重要，则 $a_{ij}=\dfrac{W_i}{W_j}=5$。反之，如果要素 j 比要素 i 明显不重要，则 $a_{ij}=\dfrac{W_i}{W_j}=\dfrac{1}{5}$。

14.3.3 AHP 的基本步骤

应用层次分析法进行决策分析的基本步骤如下。

14.3.3.1 建立递阶层次结构模型

在进行决策时，首先要明确决策的目标、准则以及待决策的方案。为此，先要分析待决策问题中所包含的要素，并按照要素间的相互关联影响以及隶属关系，将

各要素按不同层次聚集组合，形成一个多层次的结构模型，这就是递阶层次结构模型。在递阶层次结构模型中，通常第一层是目标层（最高层），它表示问题的目的、总目标；第二层是准则层，它是总目标的具体体现，是决策所要考虑的多个子目标，也是决策的具体准则；第三层、第四层……是子准则层，它们将准则更加细化；从第二层起的准则层和所有的子准则层都属于中间层；最下面的一层是方案层（最底层），它表示待选择的方案、措施、政策等。这样就形成了递阶层次结构模型，如图14－2所示。

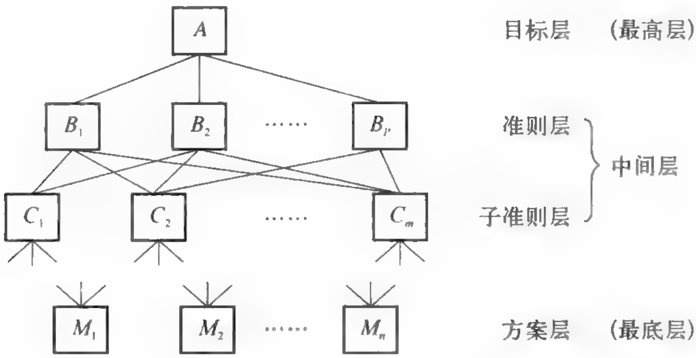


图14－2 递阶层次结构图

下面用一个例子说明如何运用AHP进行多目标决策。

【例14－3】 东方港务公司设备选择决策问题。

东方港务公司拟增添一台新设备。现有三种不同型号的设备 P_1、P_2、P_3 供选择。选择设备的主要考虑要素是功能、价格和维护，建立这一决策问题的递阶层次结构模型。

【解】 这是一个多目标决策问题，根据题意可知该决策问题的目标、准则与方案，由此可建立本问题的递阶层次结构模型，如图14－3所示。

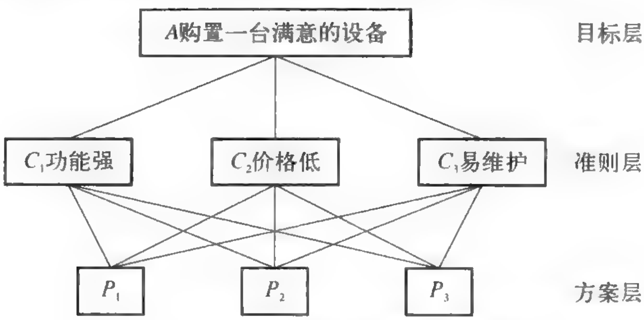


图14－3 设备购买决策的递阶层次结构图

14.3.3.2 构造比较判断矩阵

递阶层次结构建立后，第二步就是在各层要素中进行两两比较，并引入判断尺度将其量化，构造出比较判断矩阵。

设比较判断矩阵 $\boldsymbol{A}=\{a_{ij}\}$，$(i, j=1, 2, \cdots, n)$。比较判断矩阵中的元素 a_{ij} 是以上一层次某要素（比如说，要素 A）为准则，对本层次的 n 个要素（比如说，要素 $C_1, C_2, \cdots, C_n$）进行两两比较来确定的。其形式如表 14－4 所示。

表 14－4 【例 14－3】比较判断比较表

A	C_1	C_2	$\cdots$	C_j	$\cdots$	C_n
C_1	a_{11}	a_{12}	$\cdots$	a_{1j}	$\cdots$	a_{1n}
C_2	a_{21}	a_{22}	$\cdots$	a_{2j}	$\cdots$	a_{2n}
$\cdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
C_i	a_{i1}	a_{i2}	$\cdots$	a_{ij}	$\cdots$	a_{in}
$\cdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
C_n	a_{n1}	a_{n2}	$\cdots$	a_{nj}	$\cdots$	a_{nn}

其中，比较判断矩阵中的元素 a_{ij} 表示对上一层要素 A 而言，本层要素 C_i 对于 C_j 的相对重要程度。即

$$a_{ij}=\frac{W_i}{W_j} \quad (i, j=1, 2, \cdots, n)$$

式中，W_i 和 W_j 分别为准则层要素 C_i 和 C_j 的相对重要性权值。根据判断尺度将上式量化，即可得到比较判断矩阵。

【例 14－4】 构造【例 14－3】中的比较判断矩阵。

【解】 在【例 14－3】中，A 为目标层，其下一层的相关要素有三个：功能 C_1、价格 C_2、维护 C_3。现通过咨询，由专家对要素 C_1、C_2、C_3 进行两两比较，得到了如下结果："功能(C_1)"比"价格(C_2)"明显重要，比"维护(C_3)"稍微重要；"价格(C_2)"比"功能(C_1)"明显不重要，比"维护(C_3)"稍微不重要。由此可构造出比较判断矩阵 $\boldsymbol{A}_{-C}$ 如表 14－5 所示。

表 14－5 【例 14－3】比较判断矩阵表

A	C_1	C_2	C_3
C_1	1	5	3
C_2	1/5	1	1/3
C_3	1/3	3	1

所以

$$\boldsymbol{A}_{-C}=\begin{bmatrix}1 & 5 & 3\\1/5 & 1 & 1/3\\1/3 & 3 & 1\end{bmatrix}$$

然后再看下一层要素。先考虑相对于上一层次的第一个要素 C_1(功能)而言,本层相关要素 P_1、P_2、P_3 的相对重要性。通过咨询,对 P_1、P_2、P_3 进行两两比较,可构造出比较判断矩阵 $\boldsymbol{C}_{1-P}$ 如表 14-6 所示。

表 14-6 【例 14-3】比较判断矩阵表

C_1	P_1	P_2	P_3
P_1	1	1/4	2
P_2	4	1	8
P_3	1/2	1/8	1

所以

$$\boldsymbol{C}_{1-P}=\begin{bmatrix}1 & 1/4 & 2\\4 & 1 & 8\\1/2 & 1/8 & 1\end{bmatrix}$$

比较判断矩阵 $\boldsymbol{C}_{1-P}$ 表明,对于"功能 C_1"这个准则而言,方案 P_1 与 P_2 相比,其贡献的重要性介于稍微不重要与明显不重要之间;方案 P_1 与 P_3 相比,其贡献的重要性介于同等重要与稍微重要之间,等等。

同理可构造出比较判断矩阵 $\boldsymbol{C}_{2-P}$ 与 $\boldsymbol{C}_{3-P}$ 如下:

$$\boldsymbol{C}_{2-P}\begin{bmatrix}1 & 4 & 1/3\\1/4 & 1 & 1/8\\3 & 8 & 1\end{bmatrix}\qquad \boldsymbol{C}_{3-P}=\begin{bmatrix}1 & 1 & 1/3\\1 & 1 & 1/5\\3 & 5 & 1\end{bmatrix}$$

14.3.3.3 *层次单排序*

如前所述,比较判断矩阵的特征向量 W 即为各要素的相对重要性向量。因此,在得到比较判断矩阵后,接下去应计算比较判断矩阵的特征向量 W 和特征值 $\lambda_{\max}$(可以证明,该特征值是该矩阵的最大特征值)。这就是层次单排序要完成的工作。因此,层次单排序就是计算同一层次相应要素对于上一层次某要素的相对重要性排序权值,层次单排序的做法是计算各比较判断矩阵的最大特征值及其对应的特征向量。

矩阵的特征值及其对应的特征向量的估算方法有和积法、方根法、幂法等多种方法。本书介绍其中的和积法。

设有 $n\times n$ 矩阵 $\boldsymbol{A}=\{a_{ij}\}$，$(i, j=1, 2, \cdots, n)$。用和积法估算矩阵 $\boldsymbol{A}$ 的最大特征值及其对应特征向量的步骤如下：

(1) 计算比较判断矩阵 $\boldsymbol{A}$ 中每一列要素的列和 S_j：

$$S_j=\sum_{i=1}^{n}a_{ij}\quad (j=1, 2, \cdots, n) \tag{14-7}$$

(2) 将比较判断矩阵 $\boldsymbol{A}$ 中的各个要素除以该要素所在列的列和 S_j，得到一个归一化了的新矩阵 $\boldsymbol{A}_{\text{norm}}$，这里的归一化矩阵是指每一列的列和等于 1 的矩阵。设 $\boldsymbol{A}_{\text{norm}}=\{a_{ij}^{*}\}$，则有

$$a_{ij}^{*}=\frac{a_{ij}}{S_j}\quad (i, j=1, 2, \cdots, n) \tag{14-8}$$

(3) 计算新矩阵 $\boldsymbol{A}_{\text{norm}}$ 中每一行的均值 W_i，得到特征向量 $\boldsymbol{W}$。它就是 $\boldsymbol{A}$ 矩阵中各要素的层次单排序权值为

$$W_i=\frac{\sum_{i=1}^{n}a_{ij}^{*}}{n}\quad (i=1, 2, \cdots, n) \tag{14-9}$$

则 $\boldsymbol{W}=[W_1, W_2, \cdots, W_i, \cdots, W_n]^{\mathrm{T}}$ 为所求之特征向量。

(4) 计算比较判断矩阵的最大特征值为

$$\lambda_{\max}=\sum_{i=1}^{n}\frac{(\boldsymbol{AW})_i}{nW_i} \tag{14-10}$$

【例 14 - 5】 计算【例 14 - 3】中各比较判断矩阵的最大特征值及其对应的特征向量。

【解】

$$\boldsymbol{A}_{-C}=\begin{bmatrix}1 & 5 & 3\\ 1/5 & 1 & 1/3\\ 1/3 & 3 & 1\end{bmatrix}$$

由式(14 - 7)～式(14 - 10)可以进行以下计算：

(1) 计算比较判断矩阵中各列的列和为

$$S_1=\sum_{i=1}^{3}a_{i1}=1.533,\ S_2=\sum_{i=1}^{3}a_{i2}=9.000,\ S_3=\sum_{i=1}^{3}a_{i3}=4.333$$

(2) 计算归一化的新矩阵

$$a_{11}^{*}=\frac{a_{11}}{S_1}=\frac{1}{1.533}=0.652,\ a_{12}^{*}=\frac{a_{12}}{S_2}=\frac{5}{9.000}=0.556,\ a_{13}^{*}=\frac{a_{13}}{S_3}=\frac{3}{4.333}=0.692$$

$$a_{21}^{*}=\frac{a_{21}}{S_1}=\frac{1/5}{1.533}=0.130,\ a_{22}^{*}=\frac{a_{22}}{S_2}=\frac{1}{9.000}=0.111,\ a_{23}^{*}=\frac{a_{23}}{S_3}=\frac{1/3}{4.333}=0.077$$

$$a_{31}^{*}=\frac{a_{31}}{S_1}=\frac{1/3}{1.533}=0.217,\ a_{32}^{*}=\frac{a_{32}^{*}}{S_2}=\frac{3}{9.000}=0.333,\ a_{33}^{*}=\frac{a_{33}}{S_3}=\frac{1}{4.333}=0.231$$

$$\boldsymbol{A}_{\text{norm}}=\begin{bmatrix}0.652 & 0.556 & 0.692\\ 0.130 & 0.111 & 0.077\\ 0.217 & 0.333 & 0.231\end{bmatrix}$$

(3) 估计比较判断矩阵的特征向量

$$\boldsymbol{W}_1=\frac{\sum\limits_{i=1}^{3}a_{1j}^{*}}{n}=\frac{0.652+0.556+0.692}{3}=0.633$$

$$\boldsymbol{W}_2=\frac{\sum\limits_{i=1}^{3}a_{2j}^{*}}{n}=\frac{0.130+0.111+0.077}{3}=0.106$$

$$\boldsymbol{W}_3=\frac{\sum\limits_{i=1}^{3}a_{3j}^{*}}{n}=\frac{0.217+0.333+0.231}{3}=0.260$$

所以 $\boldsymbol{W}=(0.633\quad 0.106\quad 0.260)^{\mathrm{T}}$ 为判断矩阵 $\boldsymbol{A}_{-C}$ 之特征向量，亦即要素 C_1，C_2，C_3 对应于第一层 A(目标层)的相对重要性排序权值。它说明，对于总目标 A 来说，准则 C_1(功能强)、C_2(价格低)、C_3(易维护)的相对重要性权值分别为 0.633，0.106，0.260。

(4) 估算比较判断矩阵的特征根

$$\boldsymbol{A}_{-C}\boldsymbol{W}=\begin{bmatrix}1 & 5 & 3\\ 1/5 & 1 & 1/3\\ 1/3 & 3 & 1\end{bmatrix}\begin{bmatrix}0.633\\ 0.106\\ 0.260\end{bmatrix}=\begin{bmatrix}1.946\\ 0.320\\ 0.790\end{bmatrix}$$

所以 $\lambda_{\max}=\sum\limits_{i=1}^{n}\frac{(\boldsymbol{A}_{-C}\boldsymbol{W})_i}{n\boldsymbol{W}_i}=\frac{1.946}{3\times 0.633}+\frac{0.320}{3\times 0.106}+\frac{0.790}{3\times 0.260}=3.038$，

$\lambda_{\max}$ 即为判断矩阵 $\boldsymbol{A}_{-C}$ 之特征根。

同理可得：$\boldsymbol{C}_{1-P}$ 矩阵的特征向量 $\boldsymbol{W}=(0.182\quad 0.727\quad 0.091)^{\mathrm{T}}$，特征值 $\lambda_{\max}=$

3.000。可见对于准则 C_1(功能强)来说,方案 P_1、P_2、P_3 的贡献的相对重要性权值分别为 0.182,0.727,0.091。

C_{2-P}矩阵的特征向量 $W=(0.257\quad 0.074\quad 0.669)^{\mathrm{T}}$,特征值 $\lambda_{\max}=3.018$。可见对于准则 C_2(价格低)来说,方案 P_1、P_2、P_3 的贡献的相对重要性权值分别为 0.257,0.074,0.669。

C_{3-P}矩阵的特征向量 $W=(0.187\quad 0.158\quad 0.655)^{\mathrm{T}}$,特征值 $\lambda_{\max}=3.029$。可见对于准则 C_3(易维护)来说,方案 P_1、P_2、P_3 的贡献的相对重要性权值分别为 0.187,0.158,0.655。

14.3.3.4 *层次总排序*

层次单排序给出了相对于上一层次某要素,本层次相应要素的相对重要性排序权值。而最终要求出的是最底层(方案层)相对于最高层(目标层)的相对重要性排序权值,这样就可得到综合各方案在各决策准则下的优劣之后的总结果。而某层次对于最高层的相对重要性排序权值就称为该层次的层次总排序权值。层次总排序权值最大的方案就是对总目标贡献最大的方案,也就是最优方案。因此,层次总排序的目的是计算同一层次所有要素对于最高层(总目标)的相对重要性排序权值。

层次总排序是由上而下进行的。其计算过程如下:

设:在递阶层次结构模型中,最高层为 A 层;第二层为 B 层(对于第二层而言,由于它的上一层次就是最高层,所以其层次单排序权值等于层次总排序权值),B 层有 m 个要素 $B_1, B_2, \cdots, B_i, \cdots, B_m$,它们关于最高层 A 层的相对重要性排序权值分别为 $b_1, b_2, \cdots, b_i, \cdots, b_m$;$B$ 层的下一层为 C 层,设 C 层有 n 个要素 $C_1, C_2, \cdots, C_i, \cdots, C_n$,它们关于 B 层中任一要素 B_i 的层次单排序的排序权值分别为 $c_1^i, c_2^i, \cdots, c_j^i, \cdots, c_n^i$(如果 C 层中某要素 C_k 与要素 B_i 无关,则该项权值 c_k^i 为零),则 C 层中各要素对于最高层 A 层的层次总排序权值 $c_1, c_2, \cdots, c_j, \cdots, c_n$ 为

$$c_j=\sum_{i=1}^{m} b_i c_j^i \quad (j=1, 2, \cdots, n) \tag{14-11}$$

即 C 层对于总目标 A 的层次总排序权值,是以上一层次 B 的层次总排序权值为权重、对 C 层的层次单排序权值进行加权和得出的,如表 14-7 所示。

表 14-7 层次总排序

B层要素 / B层层次单排序权值 / C层层次单排序权值 / C层要素	B_1	B_2	…	B_i	…	B_m	C层层次总排序权值 c_j
	b_1	b_2	…	b_i	…	b_m	
C_1	c_1^1	c_1^2	…	c_1^i	…	c_1^m	$c_1=\sum_{i=1}^{m} b_i c_1^i$

续 表

B层要素 / B层层次单排序权值 / C层层次单排序权值 / C层要素	B_1	B_2	$\cdots$	B_i	$\cdots$	B_m	C层层次总排序权值 c_j
	b_1	b_2	$\cdots$	b_i	$\cdots$	b_m	
C_2	c_2^1	c_2^2	$\cdots$	c_2^i	$\cdots$	c_2^m	$c_2 = \sum_{i=1}^{m} b_i c_2^i$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
C_j	c_j^1	c_j^2	$\cdots$	c_j^i	$\cdots$	c_j^m	$c_j = \sum_{i=1}^{m} b_i c_j^i$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
C_n	c_n^1	c_n^2	$\cdots$	c_n^i	$\cdots$	c_n^m	$c_n = \sum_{i=1}^{m} b_i c_n^i$

如果 C 层下还有 D 层，D 层有 p 个要素 $D_1, D_2, \cdots, D_i, \cdots, D_P$，则由式(14-11)得 D 层的层次总排序权值(即 D 层对于最高层 A 层的相对重要性排序权值)为

$$d_j = \sum_{i=1}^{n} c_i d_j^i \quad (j = 1, 2, \cdots, p)$$

式中，c_i 为上一层次(C 层)的要素 $C_i (i = 1, 2, \cdots, n)$ 的层次总排序权值；d_j^i 为本层次(D 层)要素 $D_j (j = 1, 2, \cdots, p)$ 对于上一层次的要素 C_i 的层次单排序权值。若 D 层下还有 E 层、F 层、……，则用同法依次往下递推计算，最终可求出最底层(即方案层)对于总目标的总排序权值，其中总排序权值最高的方案就是最优方案。

【例 14-6】 对【例 14-3】至【例 14-5】中的问题进行层次总排序。

【解】 将单排序的结果列入表 14-7 中，得到表 14-8。

表 14-8 层次总排序结果

C层要素 / C层层次总排序权值 / P层层次总排序权值 / P层要素	C_1	C_2	C_3	P层层次总排序权值
	0.633	0.106	0.260	
P_1	0.182	0.257	0.187	$0.182 \times 0.633 + 0.257 \times 0.106 + 0.187 \times 0.260 = 0.191$
P_2	0.727	0.074	0.158	$0.727 \times 0.633 + 0.074 \times 0.106 + 0.158 \times 0.260 = 0.510$
P_3	0.091	0.669	0.655	$0.091 \times 0.633 + 0.669 \times 0.106 + 0.655 \times 0.260 = 0.299$

P 层的总排序权值为(0.191　0.510　0.299)^{T},因此,方案 P_2 为最优方案,其次是方案 P_3,而最次方案是 P_1。

14.3.3.5　一致性检验

如前所述,比较判断矩阵是由评价者通过两两比较得到的,但评价者往往很难精确地判断出比较判断矩阵中各元素的值,而只能对它们进行估计。如果在估计时偏差过大,出现严重的思维判断不一致的情况,就必须对比较判断矩阵进行修正。这种思维不一致主要表现为两种逻辑错误:

(1) 克星逻辑错误:是一种根本性的判断错误。在上述购买设备的例子中,当“功能”比“维护”重要、而“维护”又比“价格”重要时,如果认为“价格”比“功能”更重要,那就不符合逻辑了。这种自相矛盾的逻辑错误就是克星逻辑错误。

(2) 量度逻辑错误:是一种对重要程度的判断错误。当“功能”与“维护”相比的相对重要程度为 3、而“维护”与“价格”相比的相对重要程度是 5 时,如果认为“功能”与“价格”相比的相对重要程度仅为 3,那就在程度上不符合逻辑了。这种程度上的判断错误就是量度逻辑错误。

一致性检验就是检验各比较判断矩阵是否存在这两类逻辑错误,同时确定这种错误是属可接受的、还是不可接受的。只有通过一致性检验的比较判断矩阵才被认为是有效的,否则就应进行修正。

设有 $n\times n$ 比较判断矩阵 $\boldsymbol{A}$,研究发现,当 $\boldsymbol{A}$ 矩阵完全一致时,则有 $\lambda_{\max}=n$(其中 n 为 $\boldsymbol{A}$ 矩阵的阶数);当 $\boldsymbol{A}$ 矩阵稍有不一致时,$\lambda_{\max}\geqslant n$(表示 $\lambda_{\max}>n$,同时 $\lambda_{\max}$ 又接近于 n 的值);$\boldsymbol{A}$ 矩阵的不一致越大,$\lambda_{\max}$ 与 n 的差就越大。因此可以用($\lambda_{\max}-n$)来作为度量不一致性的指标。

定义一致性指标 CI 为

$$CI=\frac{\lambda_{\max}-n}{n-1} \tag{14-12}$$

定义随机一致性比率 CR 为

$$CR=\frac{CI}{RI} \tag{14-13}$$

式中,RI 为平均随机一致性指标,它是仅与比较判断矩阵的阶数有关的指标。RI 的值如表 14-9 所示。

表 14-9　平均随机一致性指标 RI 的值

矩阵阶数 n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.52	0.89	1.11	1.25	1.35	1.40	1.45	1.49

注:事实上,当矩阵阶数 $n\leqslant 2$ 时,矩阵不存在不一致问题,所以不必检验。

一致性检验的步骤如下：

(1) 由式(14-10)计算比较判断矩阵的最大特征值 λ_{max}。

(2) 由式(14-12)、式(14-13)与表 14-9 计算随机一致性比率 CR。

(3) 当 $CR \leqslant 0.1$ 时，比较判断矩阵具有满意的一致性；当 $CR > 0.1$ 时，比较判断矩阵不一致，必须进行修正。

【例 14-7】 对【例 14-3】~【例 14-6】中的比较判断矩阵进行一致性检验。

【解】 在【例 14-6】中已得出比较判断矩阵 $\boldsymbol{A}_{-C}$、$\boldsymbol{C}_{1-P}$、$\boldsymbol{C}_{2-P}$、$\boldsymbol{C}_{3-P}$的最大特征值λ_{max}，下面对这 4 个矩阵进行一致性检验。

比较判断矩阵 $\boldsymbol{A}_{-C}$的最大特征值为$\lambda_{max}=3.308$，$\boldsymbol{A}_{-C}$为 3×3 矩阵，即 $n=3$。由式(14-12)可得

$$CI=\frac{\lambda_{max}-n}{n-1}=\frac{3.308-3}{3-1}=0.019$$

查表 14-6 可得，当 $n=3$ 时，$RI=0.52$，由式(14-13)可得，$CR=\dfrac{CI}{RI}=\dfrac{0.019}{0.52}=0.037<0.1$。

所以比较判断矩阵 $\boldsymbol{A}_{-C}$具有满意的一致性。

同理可得，比较判断矩阵 $\boldsymbol{C}_{1-P}$的$CR=0.000<0.1$；$\boldsymbol{C}_{2-P}$的$CR=0.018<0.1$；$\boldsymbol{C}_{3-P}$的$CR=0.029<0.1$，所以比较判断矩阵$\boldsymbol{C}_{1-P}$、$\boldsymbol{C}_{2-P}$、$\boldsymbol{C}_{3-P}$均具有满意的一致性。

14.4 模糊决策法

14.4.1 基本概念

在现实生活中，很多概念都是模糊的。如高个子，身高达到多少即算高个子，并无明确的定义，不同的人会有不同的理解。另外，如码头装卸服务的高效、优质等的概念是模糊的；商品的舒适、美观、价格合理等概念也是模糊的。这些概念的内涵是明确的，但外延则是模糊的。正因为在现实经济生活中，很多概念都是模糊的，利用模糊数学进行决策分析的应用就越来越广，模糊决策方法正成为决策领域中一种很有实用价值的工具。

14.4.1.1 模糊集合

设 X 为一基本集，若对每个 $x\in X$，都指定一个数 $\mu_A(x)\in[0,1]$，则定义模

糊子集 $\underset{\sim}{A}$：

$$\underset{\sim}{A}=\left\{\left|\frac{\mu_{\underset{\sim}{A}}(x)}{x}\right| x\in X\right\}$$

$\mu_{\underset{\sim}{A}}(x)$称为 $\underset{\sim}{A}$ 的隶属函数；$\mu_{\underset{\sim}{A}}(x_i)$称为元素 x_i 的隶属度。

当 X 是可数集合，$X=\{x_1, x_2, \cdots, x_n\}$ 时，则

$$\underset{\sim}{A}=\sum_{i=1}^{n}\frac{\mu_{\underset{\sim}{A}}(x)}{x}$$

当 X 中元素不可数时，则记为

$$\underset{\sim}{A}=\int\frac{\mu_{\underset{\sim}{A}}(x)}{x}\mathrm{d}x \qquad x\in X$$

隶属函数 $\mu_{\underset{\sim}{A}}(x)\in[0, 1]$，即 $0\leqslant\mu_{\underset{\sim}{A}}(x)\leqslant 1$。

例如，设某 4 人 a、b、c、d 属于高个子的程度分明为 0.8，0.5，0.6，0.2，则该集合可表示成：

$$\underset{\sim}{A}=\frac{0.8}{a}+\frac{0.5}{b}+\frac{0.6}{c}+\frac{0.2}{d}$$

式中的“+”称为查德符号，表示模糊集合的元素相并列，没有相加的含义。

14.4.1.2 隶属函数的确定

实践中，隶属函数的确定有许多方法，各种方法的客观程度也不一样。下面介绍模糊统计确定隶属函数的方法。该方法是先选取一个基本集，然后取其中任一元素 x_i，再考虑此元素属于集合 $\underset{\sim}{A}$ 的可能性。例如，先确定模糊集合是高个子，然后考虑某人 a 属于高个子模糊集合的可能性，为得到量化的数据，可以邀请一些人来评判 a 是否为高个子，由于人们对高个子的认识不一样，有人会认为是，有人则会认为不是，这样可以得到

$$\mu(a)=\lim_{n\to\infty}\frac{a\in\underset{\sim}{A}\text{ 的次数}}{n}$$

这里 n 是参加评判总人数，试验次数只要充分大，$\mu(a)$就会趋向$[0, 1]$中的一个数，此数即为隶属度。

14.4.1.3 截集

模糊集合的 λ 截集是指 X 中对 $\underset{\sim}{A}$ 的隶属度不小于 λ 的一切元素组成的普通

集合，其定义如下：

对于给定的实数 $\lambda(0\leqslant\lambda\leqslant1)$，定义：

$$A_\lambda = \{x \mid \mu_A(x) \geqslant \lambda\}$$

为 A 的 λ 截集，其中 λ 称为置信水平。

14.4.2 模糊决策的应用

下面通过一个实例来介绍模糊决策的应用。

例如，某港口管理系统公司在研究产品发展方向时，有两个方案可供考虑：A 方案是生产产品型号甲，B 方案是生产产品型号乙。公司决策层对产品进行了功能分析，认为产品应具有数据分析、操作方便和界面美观三大功能，相应的功能集合为

$$X = \{X_1(\text{数据分析}),\ X_2(\text{操作方便}),\ X_3(\text{界面美观})\}$$

针对不同功能因素，由有代表性的顾客子集对这三个因素进行评述，评级域定为

$$V = \{V_1(\text{很好}),\ V_2(\text{好}),\ V_3(\text{不太好}),\ V_4(\text{不好})\}$$

对产品型号甲的"数据分析"，顾客中有30%认为"很好"，60%认为"好"，还有10%认为"不太好"，却无人认为"不好"。则对产品型号甲的"数据分析"的评价为

$$(0.3,\ 0.6,\ 0.1,\ 0)$$

相似地，可得出产品型号甲的"操作方便"和"界面美观"的评价分别为

$$(0.3,\ 0.6,\ 0.1,\ 0)$$
$$(0.4,\ 0.3,\ 0.2,\ 0.1)$$

同样，对产品型号乙的"数据分析"、"操作方便"和"界面美观"的评价为

$$(0.1,\ 0.2,\ 0.6,\ 0.1)$$
$$(0.1,\ 0.3,\ 0.5,\ 0.1)$$
$$(0.2,\ 0.2,\ 0.3,\ 0.3)$$

于是，就 A、B 两个方案，写出评价矩阵

$$\boldsymbol{R}_A = \begin{bmatrix} 0.3 & 0.6 & 0.1 & 0 \\ 0.3 & 0.6 & 0.1 & 0 \\ 0.4 & 0.3 & 0.2 & 0.1 \end{bmatrix}$$

$$R_B = \begin{bmatrix} 0.1 & 0.2 & 0.6 & 0.1 \\ 0.1 & 0.3 & 0.5 & 0.1 \\ 0.2 & 0.2 & 0.3 & 0.3 \end{bmatrix}$$

由于顾客对“数据分析”、“操作方便”和“界面美观”的要求不一样，三者有所不同，因此要考虑相应地加上不同权值，如：

数据分析给予权值 0.3；操作方便给予权值 0.3；界面美观给予权值 0.4。

以上权值满足归一化要求，即 $0.3+0.3+0.4=1$，这三个权值组成 X 上的一个模糊向量：

$$\boldsymbol{W} = (0.3, 0.3, 0.4)$$

由此得出顾客对 A、B 两个方案的综合评价：

$$\begin{aligned} \boldsymbol{B}_A &= \boldsymbol{W} \cdot \boldsymbol{R}_A \\ &= (0.3, 0.3, 0.4) \cdot \begin{bmatrix} 0.3 & 0.6 & 0.1 & 0 \\ 0.3 & 0.6 & 0.1 & 0 \\ 0.4 & 0.3 & 0.2 & 0.1 \end{bmatrix} \\ &= (0.4, 0.3, 0.2, 0.1) \end{aligned}$$

上式中是按最大最小规则求解的，如：

$$\begin{aligned} b_1 &= (0.3 \wedge 0.3) \vee (0.3 \wedge 0.3) \vee (0.4 \wedge 0.4) \\ &= 0.3 \vee 0.3 \vee 0.4 \\ &= 0.4 \end{aligned}$$

式中，“$\wedge$”表示取最小值；“$\vee$”表示取最大值。其余可类似求出。

同理

$$\begin{aligned} \boldsymbol{B}_B &= \boldsymbol{W} \cdot \boldsymbol{R}_B \\ &= (0.3, 0.3, 0.4) \cdot \begin{bmatrix} 0.1 & 0.2 & 0.6 & 0.1 \\ 0.1 & 0.3 & 0.5 & 0.1 \\ 0.2 & 0.2 & 0.3 & 0.3 \end{bmatrix} \\ &= (0.2, 0.3, 0.4, 0.3) \end{aligned}$$

因为 $0.2+0.3+0.4+0.3=1.2 \neq 1$，做归一化处理，得

$$\boldsymbol{B}_B = (0.17, 0.25, 0.33, 0.25)$$

现在，将评价结果作为自然状态概率，结合各方案的损益值，制作成模糊决策表，如表 14-10 所示。

表 14-10 模糊决策表 单位：万元

自然状态		产品评价			
		V_1	V_2	V_3	V_4
状态概率	A	0.4	0.3	0.2	0.1
	B	0.17	0.25	0.33	0.25
损益值	方案 A	1 000	800	300	−300
	方案 B	800	700	200	−200

根据表 14-10 中各自然状态概率和损益值，可计算出每一方案的期望损益值。

$$E_A = 0.4 \times 1\,000 + 0.3 \times 800 + 0.2 \times 800 + 0.1 \times (-300)$$
$$= 670(\text{万元})$$
$$E_B = 0.17 \times 800 + 0.25 \times 700 + 0.33 \times 200 + 0.25 \times (-200)$$
$$= 327(\text{万元})$$

因此，应采用方案 A，即生产产品型号甲系统。

14.5 多属性效用决策法

14.5.1 多属性效用决策的概念

本书 11.7 节所介绍的效用概率决策只涉及单目标的情况，在实际决策中，很多决策问题都是多目标决策问题，对于这类决策问题，就可以应用多属性效用决策方法。

多属性效用决策采用将目标值转化为效用值之后，再进行加权，并构成一个新的综合的单目标函数。例如，有 n 个目标，以 DV_i 表示第 i 个目标值，则该决策问题的效用值可表示为以下函数：

$$U = U(DV_1, DV_2, \cdots, DV_n)$$

通过进行综合分析并将上述效用函数进行分解，即可以将不同目标的效用值综合为单一效用值，然后就可以根据期望效用值最大原则解决多属性效用决策问题。

14.5.2 多属性效用函数

最简单的多属性效用函数是两属性效用函数。例如，某企业打算引进新设备

以提高产品质量,设备的好坏直接影响到质量的高低,但引进更好的设备通常意味着费用的提高,因此,该决策问题就不再是一个单一目标决策问题,其决策效用值是质量和费用的函数,可表示为U(质量,费用)。

辨识决策问题属性及其属性数量是建立效用函数的第一步。在建立效用函数时,应确认合理的属性及属性数量,以保证决策分析的质量。确认的属性数量太多,会增加大量的不必要的计算工作量;反之,则会影响决策分析的正确性,造成不利的决策后果。确定属性及其数量的基本原则是:确定的属性应全面可行,各属性不可再分解且没有重复,属性的数量最少。

例如,某地政府打算兴建一批平价房用于低价出租,决策分析小组原先确定的该决策问题效用函数的属性包括成本、社区服务、对本地房地产市场的影响和居民的审美价值观。后经分析,发现本地居民与外地居民的审美观存在很大差异,而该地又有许多外地居民长期居留。因此,决策分析小组最后决定将居民的审美价值观属性分解为本地居民审美价值观属性和外地居民审美价值观属性。

对于具有两个属性(以 X, Y 表示)的决策问题,定义效用函数为$U(X, Y)$。如果 X 与 Y 相互独立,则两属性效用函数可以表示为以下加性效用函数,即

$$U(X, Y) = k_1U_1(X) + k_2U_2(Y)$$

式中,k_1 和 k_2 为常数。

因此,如果两属性 X 和 Y 相互独立,我们就可以通过分别求出决策者的效用函数$U(X)$和$U(Y)$而求出该两属性效用函数$U(X, Y)$,其中,k_1 和 k_2 是两属性的相对重要性。

例如,某港口设备公司在制定价格和广告策略时面临两个决策方案,每一决策方案的实施均会引起市场占有率和投资回报率的变化,如表 14-11 所示。其决策树如图 14-4 所示。

表 14-11　某公司的决策问题

决策方案	市场占有率 X/%	投资回报率 Y/%	概　率
1	4	12	0.6
	6	6	0.4
2	5	8	0.3
	3	13	0.5
	4	8	0.2

假设通过分析,该公司决策者认为两属性——市场占有率 X 和投资回报率 Y

的效用相互独立，则该决策问题的效用函数可用加性效用函数描述，即

$$U(X, Y) = k_1U_1(X) + k_2U_2(Y)$$

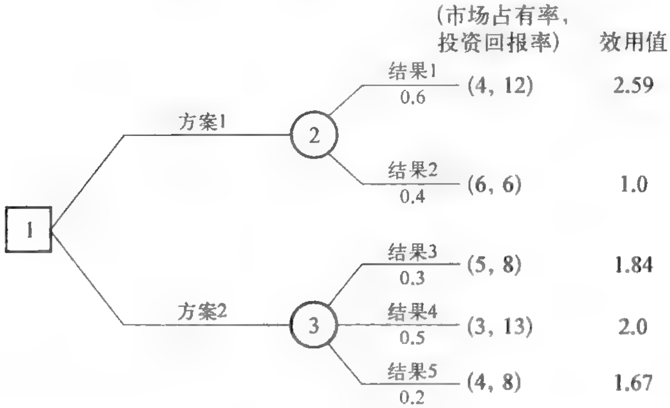


图 14-4 某公司决策问题的决策树

再假设市场占有率和投资回报率效用函数如图 14-5、图 14-6 所示。图中市场占有率和投资回报率的边界分别是市场占有率和投资回报率的最大值和最小值。

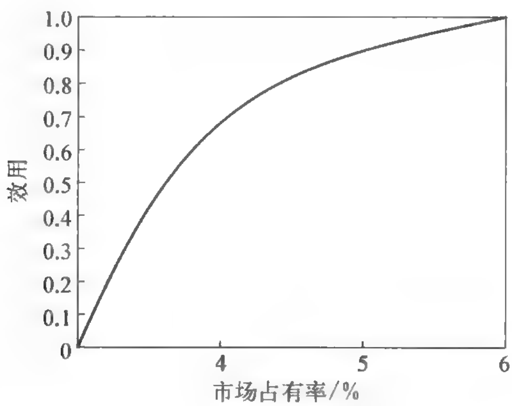


图 14-5 市场占有率的效用函数

通过进一步的分析，决策者认为投资回报率比市场占有率更重要，并赋值 $k_2 = 2k_1$。这表明投资回报率的重要程度是市场占有率的 2 倍。当然，如果该决策者的效用函数取值范围发生变化，则其权数也要调整。例如，在本例中，如果该决策者面临的最小投资回报率是 0%，最大投资回报率为 20%(区间大于目前的 8% 到 13%)，那么，即使其他条件不变，其投资回报率效用函数也会不一样，投资回报

率效用函数值的权数就会不相同。

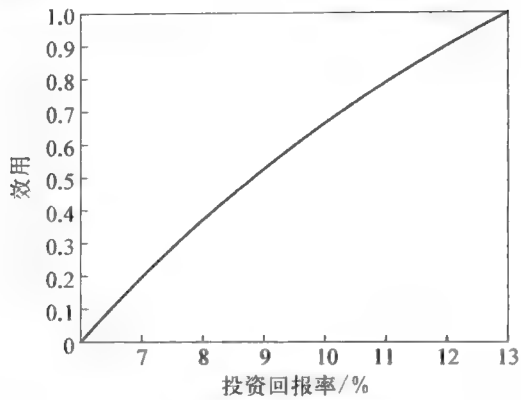


图 14-6 投资回报率的效用函数

假设经过综合分析，该决策者决定 $k_2 = 2k_1$，则可令 $k_1 = 1$，$k_2 = 2$，因此，各结果分枝的效用值可计算如下：

结果 1：$U = 1 \times 0.75 + 2 \times 0.92 = 2.59$

结果 2：$U = 1 \times 1 + 2 \times 0 = 1$

结果 3：$U = 1 \times 0.92 + 2 \times 0.46 = 1.84$

结果 4：$U = 1 \times 0 + 2 \times 1 = 2$

结果 5：$U = 1 \times 0.75 + 2 \times 0.46 = 1.67$

因此，各决策方案的期望值分别为

方案 1：$EU = 0.6 \times 2.59 + 0.4 \times 1 = 1.954$

方案 2：$EU = 0.3 \times 1.84 + 0.5 \times 2 + 0.2 \times 1.67 = 1.886$

这表明决策方案 1 优于决策方案 2。

上面介绍的两属性效用决策方法可以推广到多属性效用决策问题的决策分析中去。对于上面的两属性效用决策问题，即使我们不能假设这两个属性的效用相互独立，因而不能采用加性效用函数结构，也可以通过直接做决策者的二维效用曲面计算出各决策方案的期望效用值。如果决策问题的属性较多，则直接制作决策者的效用曲面就非常困难了。因此，假设各属性的效用相互独立，效用函数结构为加性效用函数结构形式，就能使多属性效用决策问题变得简单起来。如果各属性效用相互独立，就可以分别计算各属性的效用值，然后采取线性加权方法求出总的效用值。

很多情况下，假设各属性效用相互独立，从而采用加性效用函数结构都能够取得较高的可靠度。但在决定是否采用加性效用函数结构之前，仍需进行严谨的分

析，确保各属性效用相互独立假设的合理性。如果加性效用函数结构基本可行，但需要进行一些修正，则下式是一种可行的修正模型：

$$U(X, Y) = k_1U_1(X) + k_2U_2(Y) + k_3U_1(X)U_2(Y)$$

多目标效用函数还可以具有其他形式，如乘积形式等，它们都在实践中具有较广的应用范围。

14.5.3 加权评分法

上面的分析表明，采用加性效用函数结构能使复杂的多目标效用决策问题简单化。下面通过一个较复杂的例题，进一步介绍多目标效用决策分析。

某物流企业决定购置一辆新拖车，并且已经打算在3种设计和价格差不多的品牌中选购一辆。这3种牌号的车均符合该企业的基本要求，但该企业负责人却一时不能决定购买哪一种牌号的拖车。通过分析，该企业负责人认为其效用函数是加性的，且各属性都相互独立，这些属性包括舒适、性能、节约能源、造型、特点、引擎质量和功率，则效用函数可表示为

$$U(\text{汽车}_i) = f(\text{舒适},\text{性能},\text{节约能源},\text{造型},\text{特点},\text{引擎质量},\text{功率})$$

或

$$U(Y_i) = f(X_1, X_2, X_3, X_4, X_5, X_6, X_7)$$

该负责人通过仔细分析后认为，在上述7个属性内，可以认为各属性效用值是相互独立的。且因为要比较的3辆拖车很相似，因此，可以应用加性效用函数，即

$$U(Y_i) = k_1U_1(X_1) + k_2U_2(X_2) + \cdots + k_7U_7(X_7)$$

同时，该负责人还将这3辆拖车的不同属性进行了评分，其结果如表14-12所示，其评分标准采取从0分到10分，0分是最低分，10分是最高分。

表14-12 拖车购买决策分析

	拖车牌号 Y_1	拖车牌号 Y_2	拖车牌号 Y_3
舒适 X_1	6	9	7
性能 X_2	5	3	9
节能 X_3	9	6	7
造型 X_4	5	8	8
特点 X_5	6	7	5
引擎质量 X_6	7	6	9
功率 X_7	3	8	5

上式中的 k_i 值是各属性的相对权重，该企业负责人通过以下两两比较计算出 k_i 值。

(1) 节能的重要性是功率的 2 倍。

(2) 舒适与节能同等重要。

(3) 舒适的重要性是引擎质量的 1.5 倍。

(4) 造型与功率同等重要。

(5) 特点与引擎质量同等重要。

(6) 性能的重要性是功率的 1.5 倍。

根据上述两两比较，有

$$k_3 = 2k_6 \qquad k_1 = k_3$$
$$k_1 = 1.5k_5 \qquad k_4 = k_6$$
$$k_7 = k_5 \qquad k_2 = 1.5k_6$$

根据上述结果，可以得到一组解为

$$k_1 = 2 \qquad k_2 = 1.5$$
$$k_3 = 2 \qquad k_4 = 1$$
$$k_5 = 1.33 \qquad k_6 = 1$$
$$k_7 = 1.33$$

通过做进一步的两两比较，并为了保持各比较的一致性，最后，该负责人确定各权值如下：

$$k_1 = 1.8 \qquad k_2 = 1.5$$
$$k_3 = 2.2 \qquad k_4 = 1$$
$$k_5 = 1.33 \qquad k_6 = 1$$
$$k_7 = 1.4$$

然后，就可以计算出购买不同汽车的效用值：

$$U(Y_1) = 1.8\times 6 + 1.5\times 5 + 2.2\times 9 + 1\times 5 + 1.33\times 7 + 1\times 3 + 1.4\times 6 = 63.83$$
$$U(Y_2) = 1.8\times 9 + 1.5\times 3 + 2.2\times 6 + 1\times 8 + 1.33\times 6 + 1\times 8 + 1.4\times 7 = 67.7$$
$$U(Y_3) = 1.8\times 7 + 1.5\times 9 + 2.2\times 7 + 1\times 8 + 1.33\times 9 + 1\times 5 + 1.4\times 5 = 73.5$$

可见第 3 种牌号的拖车最优，其次为第三种牌号的拖车。

14.6 优劣系数法

14.6.1 目标权数的确定

优劣系数法是多目标决策中的常用方法之一。在计算优劣系数之前，必须首先确定各目标的权数。

例如，某码头准备新建一泊位，考虑的主要目标有建设投资、建成年限、建成后需投入流动资金、年产值、产值利税率和环境污染。表 14－13 列出了 3 个不同方案的目标值。

表 14－13 不同方案目标值

目 标	单 位	方案 1	方案 2	方案 3
建设投资	万元	1 000	860	750
建成年限	年	5	4	3
建成后需投入流动资金	万元	458	333	385
年产值	万元	2 600	1 960	2 200
产值利税率	%	12	15	12.5
环境污染		3	6	5

其中，环境污染以 1～9 之间的数字表示，数字越大，表明环境污染越轻。

从表 14－13 中可以看出，没有一个方案绝对优于其他方案，也没有一个方案的各项指标绝对劣于其他方案。一般而言，对决策者来说，各目标的重要性并不是一样的，有些目标相对重要一些，有些目标则相对次要一些，因此，需要对不同的目标给予不同的权数，下面介绍几种确定权数的方法。

1）简单编码法

简单编码法是将目标按重要性依次排序，最次要的目标定为 1，然后按自然数顺序由小到大确定权数。如有 A、B、C、D 4 个目标，依重要性排序为 B、C、A、D，则其权数分别为 4、3、2、1，将权数归一化，则权数分别是 A：0.2，B：0.4，C：0.3，D：0.1。按简单编码法计算权数方法简单，但权数差别小，欠缺合理性。

2）环比法

环比法是将各目标随机排成一行，然后按排列顺序将两个目标对比，得出环比比率再连乘，把环比比率换算为以最后一个目标为基数的定基比率，然后进行归一化处理。如有 4 个目标，其权数确定如表 14－14 所示。

表 14-14　用环比法确定目标权数

目　标	A	B	C	D	合　计
环　比	1.5	0.5	2	1	
以 D 为基环比	1.5	1	2	1	5.5
权　数	0.272 727	0.181 818	0.363 636	0.181 818	1

环比第一行的第一个数据表示 A 比 B 重要 1.5 倍，0.5 表示 B 的重要性是 C 的一半。第三行数据由第二行算出，它以 D 为基数 1，C 的重要性为 D 的 2 倍，故取值 2，B 是 C 的重要性的一半，故取值 $0.5\times2\times1=1$，以下类推。最后一行是以合计数为分母、第三行数据为分子计算出来的。

3）优序图

优序图是一个棋盘式表格，横行和纵列都是要比较的目标，每一格填上两两对比的数字，重要性可用 1，2，3，4，5 表示，数字越大，表明重要性越大。当两个目标相比时，如一个目标的重要性为 5，则另一个目标的重要性为 0；如一个目标的重要性为 4，则另一个目标的重要性为 1。假设决策者根据本原则，对本案例的目标进行两两比较后所得结果如表 14-15 所示。将各行数值加起来，即得各行的合计数，将各行合计数除以总数 75，即得各目标的权数。

表 14-15　目标两两比较值

	目标 1	目标 2	目标 3	目标 4	目标 5	目标 6	合　计	权　数
目标 1		3	4	5	3	4	19	0.253 3
目标 2	2		4	4	3	3	16	0.213 3
目标 3	1	1		4	2	2	10	0.133 3
目标 4	0	1	1		1	2	5	0.066 7
目标 5	2	2	3	4		3	14	0.186 7
目标 6	1	2	3	3	2		11	0.146 7
合　计							75	100

14.6.2　优系数和劣系数的计算

由于各项目标值的计量单位不一样，因此，在计算优劣系数之前，需要做标准化工作。标准化的公式为

$$X=\frac{99(C-B)}{A-B}+1$$

式中，A 为最好方案目标值；B 为最坏方案目标值；C 为待评价方案目标值。

例如，表 14－13 第一行建设投资目标中，方案 3 数据最好，定其为 100；方案 1 数据最差，定其为 1。则方案 2 的 X 值为

$$X=\frac{99\times(860-1\,000)}{750-1\,000}+1=56.44$$

以此类推，可得表 14－16。

表 14－16 目标与方案

目　　标	方案 1	方案 2	方案 3
建设投资（目标 1）	1	56.44	100
建成年限（目标 2）	1	50.5	100
建成后需投入流动资金（目标 3）	1	100	58.816
年产值（目标 4）	100	1	38.125
产值利税率（目标 5）	1	100	17.5
环境污染（目标 6）	1	100	67

接下来，就可以计算优系数了。所谓优系数，是指一方案优于另一方案所对应的权数之和与全部权数之和的比率。例如，将方案 1 与方案 2 相比，方案 1 只有年产值一项优于方案 2，根据表 14－15，年产值的权数为 5，因此，方案 1 对于方案 2 的优系数为 5/75＝0.066 7。同理可计算出其他优系数，如表 14－17 所示。

表 14－17 优系数计算表

	方案 1	方案 2	方案 3
方案 1		0.066 7	0.066 7
方案 2	0.933 3		0.466 7
方案 3	0.933 3	0.533 3	

优系数只反映目标优的多少，以及这些目标的重要性，而不反映目标优的程度。为了综合比较各方案的优劣，还需要计算劣系数。劣系数是通过对比两方案的优极差和劣极差来计算的。所谓优极差，是一方案与另一方案相比，对应的那些目标中优势目标数值相差最大者；所谓劣极差，是指一方案劣于另一方案的那些目标中数值相差最大者。劣系数等于劣极差除以优极差与劣极差之和。例如，方案 2 优于方案 3 的目标有目标 3、目标 5、目标 6，其差值为

$$100-58.816=41.184 \qquad 100-17.5=82.5 \qquad 100-67=33$$

82.5 为最大差值,即优极差,而方案 2 在其余目标上劣于方案 3,其差值为

$$100-56.44=43.56 \quad 100-50.5=49.5 \quad 38.125-1=37.125$$

49.5 为最大值,即为劣极差,因此,方案 2 与方案 3 相比的劣系数为

$$\frac{\text{劣极差}}{\text{劣极差}+\text{优极差}}=\frac{49.5}{49.5+82.5}=0.375$$

同理,可推得劣系数如表 14-18 所示。

表 14-18 劣系数计算表

	方案 1	方案 2	方案 3
方案 1		0.5	0.615 4
方案 2	0.5		0.375
方案 3	0.384 6	0.625	

劣系数只反映目标数的程度,不反映劣的目标数,因此在进行决策时,应综合考虑优、劣系数。

优劣系数法是根据优劣系数逐步淘汰不理想的方案。优系数的最好标准是 1,劣系数的最好标准是 0,但在实际决策时,不可能达到这一标准,因而是通过逐步降低标准而不断淘汰方案。例如,取优系数为 0.9、劣系数为 0.1 时,由表 14-17 可知,方案 2 和方案 3 与方案 1 相比的优系数都大于 0.9,因而淘汰方案 1。如取优系数 0.75,劣系数 0.39,则由表 14-18 可知,方案 2 与方案 3 相比的劣系数小于 0.39,因此,淘汰方案 3,即在本例中,方案 2 最优。

思考与练习

1. 什么是多目标决策?处理多目标决策问题有哪些准则?

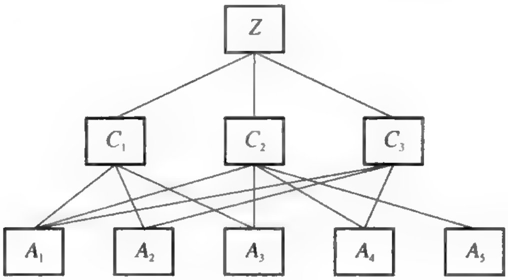
2. 解决多目标决策问题主要有哪些方法?各有什么特点?

3. 某单位经销两种货物,售出每吨甲货物可盈利 202 元,乙货物可盈利 175 元,各种货物每吨所占用的流动资金 683 元,货物经销中有 8.48%的损耗,公司负责人希望下个月能达到如下目标:

第一级要求盈利 503 万元以上;第二级要求经销甲货物 5 000 吨以上,经销乙货物 18 000 吨以上;第三级要求流动资金占用在 1 200 万元以上;第四级要求经销损耗在 1 950 吨以下。

问:单位负责人的目标是否有办法完全达到?如果无法完全达到,应如何经销才能使单位负责人最为满意?

4. 如题图 14－1 所示，设第一层总目标 Z 为“合理使用今年企业利润留成，以促进企业发展”；第二层为了衡量总目标能否实现，提出以下 3 个标准（子目标）——C_1 提高企业技术水平，C_2 改善职工物质文化生活状况，C_3 调动职工劳动积极性；第三层为了实现 C_1、C_2、C_3 目标，企业采取以下 5 个措施：A_1 办技校，A_2 扩建集体福利措施，A_3 发奖金，A_4 购置新设备，A_5 建图书馆。试用层次分析法分析该企业领导应如何使用这笔留成资金。



题图 14－1

5. 某决策问题如题表 14－1 所示，其中对科学价值、推广难易和人才培养标以 1～9 个数字，数字越大，科学价值越大，推广越容易，对人才培养越有利。试计算优系数和劣系数，并做决策分析。

题表 14－1

评价标准	D_1	D_2	D_3	D_4
经济价值/万元	50	85	70	95
科学价值	5	8	2	4
研究费用/万元	15	32	26	20
研究周期/年	1	1.5	2.5	2
推广难易	8	2	3	5
人才培养	4	8	5	3

15 管理决策模拟

在前面几章中，我们讨论了用以帮助管理人员在制定决策过程方面的一些方法和理论。几乎所有这些都属于解析技术的范畴，即用它们可直接算出数字模型决策变量的最优值。然而，在一个解析模型中的情况往往是十分复杂的，以致无法列出其表达式，或者即使将表达式列出来了，也会因其复杂性而使求解十分困难。因此，用实际系统或该系统的模型进行实验也许是唯一可行的分析方法。

本章将集中讨论具有随机的或概率性质的系统或问题，这些性质可以用蒙特卡罗模拟的形式进行分析。对大多数这类问题进行模拟，只需要纸和笔，但实际的应用则往往需要使用计算机。

15.1 模拟过程

15.1.1 引例：比尔洗车场决策问题

比尔洗车场是一个小型企业，它位于城市商业区的一条交通繁忙的单行道上。如图 15－1 所示，其设备的具体布局给作业上带来了一些限制条件，即在同一时间内只能洗刷一辆汽车，并且停车场只能停放两辆等待洗刷的汽车，沿街设立有“不准停车”“禁止停留”等指示牌，禁止停放汽车。

比尔洗车场有一个原则，这就是保证每个顾客在离开时，他的汽车是清洁的。由于采取这一原则，加上洗刷过程是由自动和人工作业综合而成的，这就使得洗刷一辆汽车所需的时间发生巨大的改变。决定洗刷时间的两个因素是：汽车的大小和汽车上堆积的污垢和灰尘多少。

企业主比尔·詹姆斯先生了解到，需要服务的顾客到达后，当他们发现停车场占满时，总是继续沿街而下，到另一个备有宽敞停车场的较大洗车场去。所以詹姆斯先生一段时间以来一直在考虑，试图对汽车洗刷作业做一些改变，以便把丢失给竞争者的一些业务争夺回来。

詹姆斯先生考虑了几个方案，诸如安装新的高速洗刷设备以及租用邻近的庭院场地等。这块场地能存放两辆等待服务的汽车。伯吉尔有限公司的业主表示愿

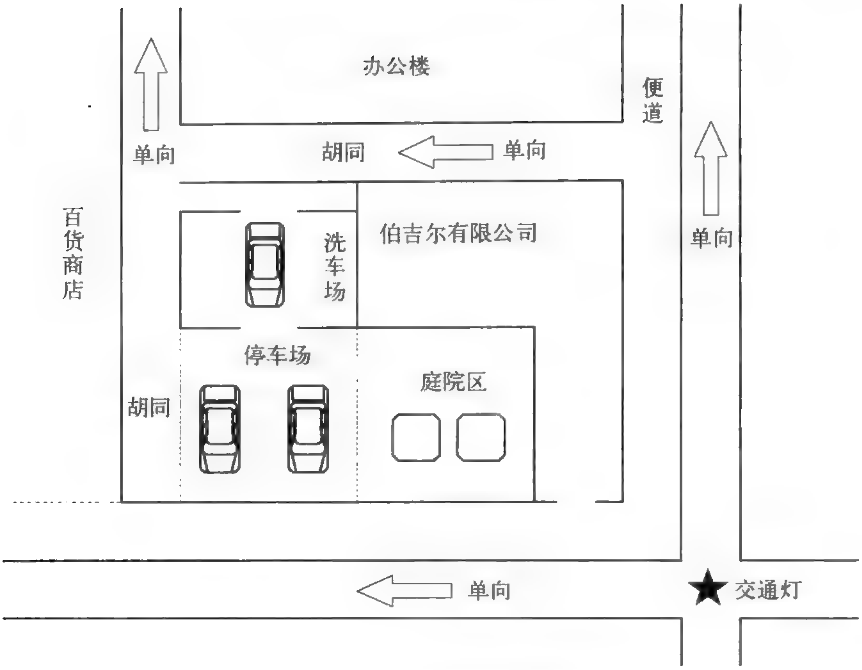


图 15-1 比尔洗车场布局图

意出租这块空地给詹姆斯先生，因为他的顾客很少有人愿意冒着噪声和烟气的干扰去使用放在那里的野餐长凳。

但是詹姆斯先生在他确信哪些办法可以增加利润以前，是不愿做出什么改变的。他决定在许多可供选择的方案中，选出几个来评价其效果。詹姆斯先生在大学的课程中，学会了模拟的初步知识，他认为可以通过汽车洗刷业务的模拟来评价这些方案。

像在比尔洗车场出现的排队或等待问题，代表了一类常常借助于模拟模型进行分析的问题。在现实生活中，许多复杂的排队问题不可能通过任何可用的排队模型来恰当地处理。因此，用实际系统或该系统的模型进行实验也许是唯一可行的分析方法。

15.1.2 模拟概述

模拟是试图重现某一系统或活动的作业情况，而不必实地去建造并运转一个系统，那样会造成巨大的花费。它是一种试验的方法。分析人员首先对关心的系统或问题构造出一个模型，然后不是按照传统的观念去求解模型，而是在各种不同的作业条件的组合下，去运行这一模型。从这些模拟运行所得出的结果，可用以对

所提出方案的优缺点进行评价。实质上,模拟给管理人员提供了一个进行试验的实验室,在这里一个问题可以在周密控制的条件下进行研究。

模拟可用来分析复杂的问题,同时其基本原理可为那些仅具备有限的数学基础的人员所理解和应用。模拟还有一些其他的优点,我们将在后面加以讨论。综合这些优点,就使模拟成为应用最广的管理科学技术之一。

模拟过程基本上可分为 5 个步骤:

(1) 对所研究的问题或系统建立模型。

(2) 为模型规定一组作业特性。

(3) 设计与进行一系列试验,以供了解模型在指定条件下的运行情况。

(4) 分析模型的运行结果,并对实际系统在类似作业条件下将如何运行进行推断。

(5) 通常,还要规定一组不同的作业特性,对模型重新进行试验,即返回到第二个步骤。这一计算过程,能对不同作业条件下的结果进行比较,以便帮助我们选择实际系统的作业条件。

15.2 建立模型

模拟过程中第一步是为所研究的过程或系统建立模型。通常这一模型是由实际中抽象出来的,模型应该足够的复杂以包含实际系统的各项重要特征。一个模拟模型通常比解析方法求解的模型所要求包含的内容更为详细,但模型应尽可能地简单。不必要的详细只能增加进行模拟所需要的时间与费用,并使输出数据的分析复杂化。

在模型构造方面,分析人员必须清晰地指出系统的各个部分,以及这些部分之间的关系。这样就可以使分析人员能把展开的模拟模型细分为若干段,以简化其结构。这样构造的模拟模型就可以预测实际系统各部分之间相互作用的结果。

15.2.1 流程图

把所研究的系统画成简图或表示流程图总是有帮助的。比尔洗车场的流程图如图 15-2 所示。第一个重要事件是汽车到达洗刷设施,然后,提出一个问题:“洗车场是否空闲?”,如果回答“是”,则汽车直接进入洗车场,如果回答“否”,则提另一个问题:“停车场是否可用?”;如果对第二个问题的回答“否”,则汽车的到达发生“障碍”或拒绝加入排队,假如回答“是”,则汽车加入等待线,并且等待进入洗刷设备。一旦汽车进入洗刷设备,即进行洗刷,然后退出系统。

图 15-2 所示的模型十分简单,它包括了詹姆斯先生所面临的问题的主要特征,因此,它可以作为以后几个步骤的基础。

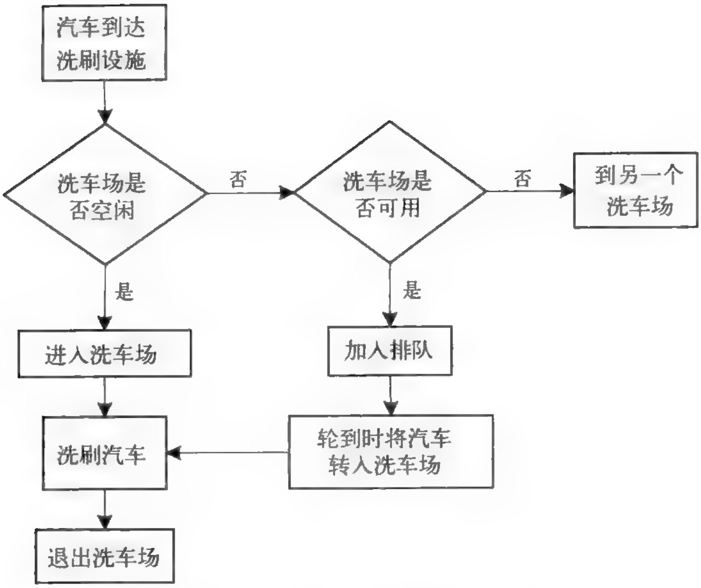


图 15－2　比尔汽车洗车场的图示

15.2.2　确定作业的特征

图 15－2 所示的比尔洗车场的流程图指出了几个必须确定的作业特征。我们所关心的事情主要包括：需要服务的汽车的到达方式、可供等待服务的汽车停留的车位数目以及洗刷服务过程本身的特性和持续时间等。

描述模拟到达过程通常所用的方法是确定各种到达间隔时间(连续到达之间的时间)的概率分布。詹姆斯先生让他的管理人员观察并记录了大量的连续到达之间的时间，确定了出现不同到达间隔时间的相对频率，并用它建立了概率分布表，如表 15－1 所示。

表 15－1　比尔洗车场到达间隔时间

间隔时间 T_i/min	概率 P_i	期望值 $T_i P_i$	间隔时间 T_i/min	概率 P_i	期望值 $T_i P_i$
0	0.06	0.00	6	0.10	0.60
1	0.10	0.10	7	0.08	0.56
2	0.13	0.26	8	0.05	0.40
3	0.17	0.51	9	0.03	0.27
4	0.15	0.60	10	0.01	0.10
5	0.12	0.60	合计	1.00	4.00

这一分布是经验分布，因为它是由观测数据直接决定的。到达间隔时间在0～10 min之间变化，其平均间隔时间为4 min。按这一平均数计算，每小时将到达15辆汽车(60 min除以平均间隔时间4 min)。

注意，这一离散分布是由我们将每一观测数据化整到最接近的分钟数而得到的。因此，间隔时间0代表了连续到达的间隔时间在30 s及以下。必须注意所采用的这种化整方法，应保证不使模拟的结果产生偏差。我们可以采用较小的时间增量或者用连续分布来表示到达间隔时间。然而，一般的研究方法不论在哪一种情况下都是相同的，而且用离散分布更容易说明问题。

建立为顾客服务所消耗的时间的概率分布可用以描述服务时间。假定詹姆斯先生的管理人员也记录了汽车从开始洗刷直到它离开这一系统之间的消耗时间总数，这一工作必须对大量汽车反复地进行，从测得的洗刷时间(化整为最接近的分钟数)的相对频率，用来建立由表15-2给出的概率分布。

表15-2 比尔洗车场为Ⅰ型汽车和Ⅱ型汽车服务的时间

服务时间 T_i/min	Ⅰ型汽车(45%)		服务时间 T_i/min	Ⅱ型汽车(55%)	
	概率 P_i	期望值 $T_i P_i$		概率 P_i	期望值 $T_i P_i$
2	0.60	1.2	4	0.20	0.8
3	0.40	1.2	5	0.40	2.0
	期望服务时间=2.4		6	0.25	1.5
			7	0.10	0.7
			8	0.05	0.4
				期望服务时间=5.4	

服务时间分两种不同型号的汽车给出。Ⅰ型汽车是小型汽车，Ⅱ型汽车是大型汽车。当然，两种分布可以合并成为一个单一的分布，但它们的差距应该保留下来，因为詹姆斯先生可能需要确定洗刷方式和定价标准，因而区别进入洗刷设施的汽车型号是有帮助的。

取样的汽车包括45%的小型汽车(Ⅰ型)和55%的大型汽车(Ⅱ型)。Ⅰ型汽车的平均服务时间为2.4 min，而Ⅱ型汽车的平均洗刷时间为5.4 min，洗刷每一辆汽车所需总平均服务时间为4.05 min，即$0.45\times2.4+0.55\times5.4$。因此，需要超过了能力，汽车洗车场必然会损失一些业务。在业务持续繁忙的期间，每小时平均洗刷汽车14.8辆(60 min除以每车4.05 min)，这一数字略低于每小时15辆汽车的平均到达率。

初始模拟是按等待线的容量为两辆汽车进行试验的，并根据先来先服务的原则洗刷汽车。

15.2.3 模拟模型的验证

从模拟得出实际系统在一定作业条件的范围内所具有的特性的结论以前，我们必须尽一切努力以验证模型的精确性，这在模拟过程中是最困难和具有挑战性的阶段之一。假如模型代表一个现在已有的系统，方法之一是在现实环境的条件下进行模拟，这样模型的输出可与实际作业系统中所收集起来的数据相比较，假如两者合理的一致，则可建立模型的可信度。

15.3 随机数的产生

为了试验一个模拟模型，需要想出一种方法产生输入信息以使模型得以运行。在许多情况下，输入信息是一些随机变量，也就是说，在某一给定的时刻，它们可选取值域内的任一数值，就要使用到随机数。所谓随机数是指在两个极端值之间均匀分布的数字。有关随机数的产生方法主要有 3 种，我们结合下面报童问题的例子来进行介绍。

【例 15－1】 假设有一个报童希望能够找到一个方法改善他订报的法则，以获得最多的利润。但是现在他不知道未来每天究竟会有多少顾客会买他的报纸，现在只根据以往经验可知顾客每天的需求量及其对应的相对次数如表 15－3 所示。

表 15－3 每天报纸需求量及其相对次数

需求量	150	160	170	180	190	200	
相对次数	1/12	2/12	4/12	3/12	1/12	1/12	12/12

现在可以用模拟的方法来产生随机数。

15.3.1 手工法(manual method)

此法采用抽签或抽球的方式，根据所要模拟的概率签条或者不同颜色的球分别代表不同的需求量来进行模拟，假设我们要两位数的模拟，可以在圆筒中放置 120 个签，其中 10 个写上 150，20 个写上 160 ，40 个写上 170，30 个写上 180，10 个写上 190 和 10 个写上 200。将所有签条搅和后，从中抽取一个，其所标明的数字即为当天的需求量，然后将此签放回，再抽第二次，即得第二天的需求量，以此类推。假设连续抽取 10 天，可能的次序如表 15－4 所示。

表 15-4 手工法产生的每天需求量

天数	1	2	3	4	5	6	7	8	9	10
需求量	180	170	200	180	170	160	170	150	170	180

如果我们把这 10 天的需求量做成次数分布表，求出次数比的概率分布，与表 15-3 所要模拟的概率分布比较，它们之间很可能有差异，但是如果模拟的日数增加，则模拟的结果必然会趋近于表 15-3 的概率分布，因为签条的安排配置方式就是按照所要模拟的对象而设计的，如果每个签条有同等被抽出可能，其结果应该相同。

15.3.2 随机数表(random number table)

目前统计上已编有多种随机数表，其中普遍使用的是兰德公司(Rand Corporation)所发展出来的随机数表，我们在这里给出一个两位数的随机数表(见表 15-5)。现在以表 15-3 说明如何以随机数表产生取自该概率分布的随机观测值。设从 00～99 一百个数字中，若出现 00～07 八个数字则表示需求量为 150 份，若出现 08～23 十六个数字则表示需求量为 160 份，……若出现 88～95 八个数字则表示需求量为 200 份。倘若出现 96～99 中任一数字，则弃而不用。依次可知，出现 150 份需求的概率为 $8/96=1/12$，出现 160 份需求的概率为 $16/96=2/12$，……出现 200 份需求的概率为 $8/96=1/12$，皆一一对应与其相对次数，于是可得表 15-6。

表 15-5 两位数字的随机数表

33	24	52	87	13	31	14	53	65	35	02	76	7	62	93	67	23	93	42	16
50	72	85	56	18	51	49	20	94	94	06	43	9	07	51	70	88	54	35	75
13	19	79	96	61	23	74	91	76	76	17	84	97	48	48	80	77	34	90	29
82	20	86	44	47	63	04	98	43	43	32	33	63	46	79	66	60	33	70	97
59	91	72	29	60	07	04	83	73	73	70	95	41	55	44	20	07	28	93	97
30	88	20	80	29	98	80	68	52	52	55	91	46	92	56	92	57	78	33	63
24	95	12	56	03	08	83	06	15	15	62	57	59	41	90	31	90	56	73	29
02	38	21	96	23	78	87	31	54	54	30	14	18	10	08	79	38	98	35	86
15	41	99	86	67	63	04	76	94	94	06	97	79	66	68	03	21	95	38	21
38	51	58	80	61	85	21	26	52	52	45	33	21	56	21	88	83	65	29	48
12	08	04	33	62	78	49	42	61	61	15	22	03	98	17	69	41	82	45	92
85	23	36	43	13	37	21	50	09	09	96	91	17	31	62	90	86	94	58	31
92	55	01	88	68	65	65	08	96	96	94	86	06	29	28	30	66	52	76	36

续 表

79	27	84	90	59	40	21	83	87	87	88	93	63	34	10	54	73	56	35	67
59	80	13	96	77	11	15	89	47	47	28	61	22	45	41	66	88	09	96	62
11	26	06	05	73	01	49	45	69	69	61	47	39	71	66	12	28	67	52	16
97	54	15	28	63	60	82	89	02	02	10	73	20	47	08	55	52	99	90	86
39	47	73	96	60	18	56	00	98	98	17	68	62	06	10	66	60	11	90	25
71	14	64	64	28	82	47	36	87	87	48	60	37	90	15	80	54	43	61	93
16	59	96	95	85	66	88	02	37	37	88	98	85	02	90	02	13	60	60	96
77	87	27	72	76	79	15	68	23	23	46	80	88	57	41	10	33	15	47	86
78	15	18	02	65	23	14	08	54	54	62	61	13	23	43	01	23	29	77	62
40	37	69	32	79	84	37	50	76	76	23	54	60	23	72	17	56	86	67	11
83	43	17	25	54	99	29	15	13	13	64	94	50	87	51	76	82	18	73	00
28	71	96	61	27	84	81	27	99	99	34	53	34	21	74	53	95	01	11	35

表 15-6 用随机数法产生报纸需求量

需 求 量	相对次数	两位整数	概 率
150	1/12	00～07	8/96
160	1/12	08～23	16/96
170	4/12	24～55	32/96
180	3/12	56～79	24/96
190	1/12	80～87	8/96
200	1/12	88～95	8/96
	12/12		96/96

15.3.3 计算机产生随机数

由于前面的产生方法颇为费时，在大规模的模拟时无法采用，近年来计算机的供应商们给出了很多套应用程序模块，只需要读进一个随机数的**种子(seed)**，就可以由电脑事先设定的程序连续产生 0～1 之间的随机数。产生随机数的程序本身有多种计算方法，而最为广泛使用的是**同余法(congruential method)**，有加性、乘性和混合性 3 种，其中以**混合性同余法(mixed congruential method)**最为常见，特予说明如下：

设 x_0 为初始随机数(称为种子)，可为任何正整数，则后一随机数的产生根据前一个随机数计算而来，其递推公式为

$$x_{n+1} = (ax_n + c)(\bmod m)$$

式中，a、c 与 m 皆为正整数，且 $a<m$，$c<m$。此式表示 x_{n+1} 为 (ax_n+c) 除以 m 的余数，x_{n+1} 的可能值为 0，1，…，$m-1$，这样最多可产生 m 个不同的随机数。往往对于 a 与 c 必须慎重选择，使数列 x_1，x_2，…，x_m 包含 m 个可能的随机数且每一数恰好出现一次。就二进制电子计算机而言，若其一字位(word)含有 b 个二元字节(bit)，则通常令 $m=2^b$，表示其在字位范围内非负整数的总个数。

一般而言，由 $m=2^b$ 所产生的随机数，其位数较大，而应用模拟上所需要的随机数的位数可能较少。例如在某随机数表中每一随机数为 5 位数，而在模拟时，若只需要 3 位数，则每一随机数只取其中连续 3 位，其余的不用。惯例之一是用最前 3 位或最后 3 位。简单起见，我们使用两位数的随机数表，如表 15-5 所示。

15.4 蒙特卡罗法

蒙特卡罗法是一种随机取样的技巧，也就是一种依预定的概率分布而产生随机数的程序。蒙特卡罗模拟法是由纽曼命名的，在第二次世界大战期间，在洛沙拉莫斯科学实验室(Los Alamos Scientific Laboratory)的物理学家被中子的动态所困惑，于是数学家纽曼和阿拉穆提议一种求解方法，逐步将个别事项依某种不规则的出现概率而求得近似的答案。因为在洛沙拉莫斯的工作要保持高度机密，纽曼将此种依人为的不规则概率分布而随机取样的求解程序予以**暗号(code name)**蒙特卡罗(Monte Carlo)。因此，以后将此类依照某种预定概率分布而随机取样的模拟技巧称为蒙特卡罗。

由以上的说明可知，蒙特卡罗法常用于某种视概率而定的问题的求解，对此种问题难以从事实际试验，也无法运用准确的数学公式加以分析，而且也不能从真实的母体中抽取样本。因而，运用蒙特卡罗法时，首先需要确定该变量的概率分布，然后应用随机数从此概率分布中获取有关变量特征的资料。所以事实上蒙特卡罗模拟是运用随机数产生与实际经验具有相同分布特征的一些变量，从而计算模拟问题。

15.4.1 随机数的利用

蒙特卡罗方法在概念上是简单的，且可以用一种简单的解释引出。假定抛一枚硬币，以确定两名选手中谁先发球。给定硬币的正面以及一个公正的抛掷机会，则每位选手取得先发球的机会是相等的。假定硬币的正面认为是选手 1 的一次“成功”，硬币的背面认为是选手 2 的一次“成功”。首先取得 5 次成功的选手，判为竞赛中先发球者。

实际上，与其抛掷一枚硬币，不如用表 15-5 中给出的随机数去模拟硬币的抛

数，以选择先发球的选手。取得 01 到 50 的数，表示硬币抛成正面，而取得 51 到 100 的数，表示抛成反面。我们随机地在任一点进入表 15－5，并按一预先确定的模式去选取随机数。假定第一个被选取的数是表中右上角的 16，而且我们预先确定的模式是逐个选取该列从上到下的数，那么，所选取得随机数列为 16、75、29、97 等。将这些数列入表 15－7 中。

表 15－7 还给出了与随机数相关联的事件。第一个数为 16，所关联的事件为硬币的正面，它意味着选手 1 的一次成功。同理，75 关联着硬币的背面，意味着选手 2 的一次成功。继续这一过程，直到一选手取得 5 次成功为止。

表 15－7 模拟抛掷硬币游戏

随机数	关联事件	获胜者
16	正面	1
75	背面	2
29	正面	1
97	背面	2
97	背面	2
63	背面	2
29	正面	1
86	背面	2

方法的基本思想是制定随机数序列，使得在给定的序列中取得一个数的概率正好等于该序列所关联的事件的概率。取得数 01～50 的概率为 0.5，它正好等于抛掷硬币得到正面的概率。一些补充例子将有助于弄清楚这一过程。

研究从 52 张洗匀的扑克牌中抽出一张的模拟。在每次抽出之后，将牌放回，并重新洗匀。抽得某一种花色的概率为 0.25。因此，可用随机数 01～25 表示黑桃，随机数列 26～50、51～75、76～100 则分别表示抽得红心、方块及梅花。

在有些情况中，需要将某些我们制定的随机数跳过去不用。假如我们模拟一粒骰子的重复抛掷，序列 01～10 可与抛成 1 点相关联，同样，序列 11～20、21～30、31～40、41～50、51～60 可以分别与抛成 2、3、4、5、6 点相关联。每当我们取得一个大于 60 的数时，可以简单地将它忽略不计，接着再取下一个随机数。

最后，应该看到，可能需要大于两位数的随机数。假定我们是在模拟人出生的月份，这至少需要 3 位随机数，序列 01～30 与 1 月相关联，32～59 与 2 月相关联，以此类推，直到 335～365 与 12 月相关联。更大的随机数可以将表 15－5 的两列或多列相组合而获得。

15.4.2 到达时间与服务时间的产生

上面所描述的过程,可以用来产生模拟洗车场所需要的输入数。在表 15－8 中,对需要洗刷汽车的各种可能到达间隔时间,指定了一系列随机数。分配给某一给定事件的随机数个数,与随机数总数(它代表各种结果的总分布)的比值,等于该给定事件出现的概率。在表 15－8 中,到达间隔时间为 3 min 的概率为 0.17,故分配随机数 30～46 代表 3 min 的到达间隔时间。因为 30～46 有 17 个数,占总分配数的 100 个数的 17%,因此,取得该范围内的数的概率为 0.17。

表 15－8　比尔洗车场中到达间隔时间随机数的分配

到达时间/min	概　率	随机数
0	0.06	01～6
1	0.10	07～16
2	0.13	17～29
3	0.17	30～46
4	0.15	47～61
5	0.12	62～73
6	0.10	74～83
7	0.08	84～91
8	0.05	92～96
9	0.03	97～99
10	0.01	00
	总计　1.00	

按照表 15－8,随机数序列 36、47、82 及 13 将依次代表 3、4、6 及 1 min 的到达间隔时间。抽取随机数的过程可以无限地继续下去,需要多少输入量就去多少,直至完成模拟。

在产生到达汽车的服务时间之前,先要确定到达的是小型汽车还是大型汽车。我们也可以用随机数来确定,因为Ⅰ型车到达的概率是 0.45,故用随机数序列 01～45 代表这种车型,其余的数 46～00 代表Ⅱ型车,这一信息在表 15－9 中给出。

服务时间长度也能通过蒙特卡罗法产生。表 15－10 表示的是两种服务时间分布的随机数分配。Ⅰ型车需要 2 min 洗刷时间的概率是 0.6,需要 3 min 服务时间的概率是 0.4。因此,对Ⅰ型车,随机数列 01～60 分配给代表 2 min 的洗刷时间,其余 40 个数 61～00,分配给代表 3 min 的洗刷时间。

例如,假定一辆汽车按到达要求洗刷。从随机数表中相邻的两个位置上抽读

表 15-9 比尔洗车场对不同车型的随机数分配

车　　型	概　　率	随 机 数
小型(Ⅰ型)	0.45	01～45
大型(Ⅱ型)	0.55	46～00
	1.00	

表 15-10 比尔洗车场对不同车型的随机数分配

服务时间/min	Ⅰ型汽车(小型)		服务时间/min	Ⅱ型汽车(大型)	
	概　率	随机数		概　率	随机数
2	0.60	01～60	4	0.20	01～20
3	0.40	61～00	5	0.40	21～60
			6	0.25	61～85
			7	0.10	86～95
			8	0.05	96～00

的数是 96 及 23，其中第一个数用来代表车型。因为 96 是在 46～00 的范围内，故它是Ⅱ型车；再参照表 15-10 中Ⅱ型车随机数的分配，由第二个数可以确定出汽车的洗刷时间，随机数 23 确定的洗刷时间为 5 min。

15.5 模拟模型的完成与分析

至此，对于开始模拟运行的初始基础工作，大部分是完成了。现在，我们准备产生输入量，并按照模型中所规定的关系去处理它们。首先要运行的是模拟现行的汽车洗刷系统。

必须研究出一种方法来处理输入量，记录运行结果，追踪系统运行的状态等。在大型模拟研究中，这是通过计算机模拟模型来完成的。对于汽车洗车场的例子，这种模拟用一支笔几张模拟工作记录表就可以完成。

15.5.1 输入数据

对于 20 个不同顾客，为确定其到达间隔时间、车型及服务时间所用的随机数，列于表 15-11 中。这 20 个数就作为比尔洗车场的初始模拟到达者。表 15-11 中第 2 列的随机数是从表 15-5 的第 1 列中取得的。它们与表 15-9 中随机数的分配结合使用，产生出第 3 列中的到达间隔时间。表 15-11 中第 5 列及第 7 列中的

表 15-11 比尔洗车场到达及服务时间的模拟 时间单位：min

(1) 到达序号	(2) 随机数	(3) 到达时间间隔	(4) 到达时刻	(5) 随机数	(6) 到达车型	(7) 随机数	(8) 服务时间
1	33	3	3	24	Ⅰ	52	2
2	50	4	7	72	Ⅱ	85	6
3	13	1	8	19	Ⅰ	79	3
4	82	6	14	20	Ⅰ	86	3
5	59	4	18	91	Ⅱ	72	6
6	30	3	21	88	Ⅱ	20	4
7	24	2	23	95	Ⅱ	12	4
8	02	0	23	38	Ⅰ	21	3
9	15	1	24	41	Ⅰ	99	3
10	38	3	27	51	Ⅱ	58	5
11	12	1	28	08	Ⅰ	04	2
12	85	7	35	23	Ⅰ	36	2
13	92	8	43	55	Ⅱ	01	4
14	79	6	49	27	Ⅰ	84	3
15	59	4	53	80	Ⅱ	13	4
16	11	1	54	26	Ⅰ	06	2
17	97	9	63	54	Ⅱ	15	4
18	39	3	66	47	Ⅱ	73	6
19	71	5	71	14	Ⅰ	64	3
20	16	1	72	59	Ⅱ	96	8

注：到达者假定出现在该分钟末。

随机数，由表 15-5 的第 2 列及第 3 列取得，它们与表 15-10 一起使用，以分别确定第 6 列和第 8 列中所表示的车型及服务时间。

表 15-11 第 4 列中给出的到达时刻，可由到达间隔时间计算出来。假定在时刻 0 汽车洗车场开始营业，第一个顾客到达间隔时间或者到达时刻，可以理解为从开始营业的时间到第一个顾客出现的时刻所经过的一段时间。对第一个顾客所产生的到达间隔时间是 3 min，因此第一个“到达”出现在时刻 3。

第二次到达所确定的到达间隔时间为 4 min，加上第一个顾客的到达时刻，就得到第二次的时刻为 7。对于顾客 3 和 4，到达间隔时间为 1 min 和 6 min，由此可以得到这些顾客到达的时刻为 8 和 14。其余顾客的到达时刻，同样可以用每一顾客的到达间隔时间加上前一顾客的到达时刻得到。

因为对于顾客 8 所产生的到达间隔时间为 0，所以顾客 7 和 8 的到达时刻是相同的。应该注意的是所有到达者均假定在该分钟末出现。

【例 15-2】 (罐子游戏)该游戏是对两个没有标记的罐子作随机选择,然后从所选的罐中取球。根据取得之球的颜色,给予奖赏。一个罐子中有 5 个红球,2 个白球,3 个蓝球。第二个罐子中有 4 个红球,4 个白球及 2 个蓝球:试建立必需的随机数分配,并利用它们来模拟四次游戏。

【解】 列表如表 15-12~表 15-13 所示。

表 15-12 选择罐子的随机数分配

罐 子	概 率	随机数
1	0.5	01~50
2	0.5	51~00

表 15-13 取球的随机数分配

罐子 1			罐子 2		
颜色	概率	随机数	颜色	概率	随机数
红	0.5	01~50	红	0.4	01~40
白	0.2	51~70	白	0.4	41~80
蓝	0.3	71~00	蓝	0.2	81~00

从表 15-5 中的第一列取出前 8 个数,并将它们分成 4 对两位数字的随机数。每对数中的第一个数用来决定选择哪一个罐子;第 2 个数用来决定取出的球的颜色。

33,50:罐子 1,红球

13,82:罐子 1,蓝球

59,30:罐子 2,红球

24,02:罐子 1,红球

15.5.2 模拟工作记录表

表 15-11 详细列出了洗车场首先到达的 20 辆车中每一辆的确切到达时刻和所用的服务时间。该表提供了模拟所需的全部输入数据。如表 15-14 所示的工作记录表称为模拟工作记录表,它用来记录系统的运行进程。我们将讨论模拟工作记录表中每一列的内容,并解释每项是如何计算的。但是,必须记住,这个工作记录表并没有标准的格式。表中每一列的安排,必须根据所模拟的问题特地加以设计。

表 15-14 等待线最大容量为 2 辆汽车的情况下，比尔洗车场 20 次汽车到达的模拟

时间单位：min

(1)到达号	(2)达到时刻	(3)到达类型	(4)服务时间	(5)到达状态	(6)是否加入系统	(7)*开始服务时刻	(8)*结束服务时刻	(9)排队时间	(10)空闲时间
1	3	Ⅰ	2	0	是	4	5	0	3
2	7	Ⅱ	6	0	是	8	13	0	2
3	8	Ⅰ	3	1	是	14	16	5	0
4	14	Ⅰ	3	1	是	17	19	2	0
5	18	Ⅱ	6	1	是	20	25	1	0
6	21	Ⅱ	4	1	是	26	29	4	0
7	23	Ⅱ	4	2	是	30	33	6	0
8	23	Ⅰ	2	3	否	—	—	—	—
9	24	Ⅰ	3	3	否	—	—		—
10	27	Ⅱ	5	2	是	34	38	6	0
11	28	Ⅰ	2	3	否	—	—	—	—
12	35	Ⅰ	2	1	是	39	40	3	0
13	43	Ⅱ	4	0	是	44	47	0	3
14	49	Ⅰ	3	0	是	50	52	0	2
15	53	Ⅱ	4	0	是	54	57	0	1
16	54	Ⅰ	2	1	是	58	59	3	0
17	63	Ⅱ	4	0	是	64	67	0	4
18	66	Ⅱ	6	1	是	68	73	1	0
19	71	Ⅰ	3	1	是	74	76	2	0
20	72	Ⅱ	8	2	是	77	84	4	0

注：假定服务作业在该分钟初开始，服务结束在该分钟末。

表 15-14 的第 2、3 和 4 列包括了每辆汽车的到达时刻、汽车类型和服务时间的有关信息。这些信息是直接从表 15-11 中取得的。表 15-14 的第 5 列提供了新的到达者到达时系统的状态。系统状态就是该系统中的汽车数量，包括正在洗刷的汽车和在等待线上的汽车。这个汽车数目，可以通过确定有几个服务结束时间大于或等于该到达时刻来计算。例如，我们研究一下第 6 个到达者在时刻 21 到达时系统的情况。在表的第 8 列中，有一个服务结束时间大于 21(25，对第 5 个到达者)，因此，第 6 辆汽车的到达状态就是 1。同样，第 8 辆汽车的到达状态是 3，因为有三辆先到达的汽车(5、6 和 7)，其服务结束时间都大于 23，即大于第 8 辆汽车的到达时刻。

第 6 列仅表示到达的汽车是加入本系统还是转到另一汽车洗刷场。如果到达状态是 2 或小于 2，则在第 6 列内填上“是”，表示到达者决定在比尔汽车洗车场等

待服务。如果到达状态是3,表示两个停车车位正被占用,则第6列内填入“否”,表示该到达者不加入本系统。

第7列表示何时开始服务。如果到达状态是0,则服务在汽车到达后立即开始。例如,第1辆汽车在第3分钟末到达,第4分钟初就开始服务。如果系统状态大于0,则在以前到达者中最晚的服务结束时刻之后立即开始服务。第6辆汽车在第21分钟到达,其到达状态是1,在以前到达者中,最晚的服务结束时刻是第25分钟,因此,服务开始时刻是第26分钟。

第8列表示服务结束时刻。这个时刻是服务开始时刻加上服务时间得到的。应该记住,服务开始时刻是由该分钟初开始计算,而服务结束时刻按该分钟末计算。因此,第1辆到达汽车的服务开始时刻是第4分钟初,服务时间是2 min,表示服务结束时刻是在第5分钟末。

第9列给出了排队时间。它表示某一汽车在等待线上的时间。它可以从该汽车的“服务开始时刻”和“到达时刻”中求得。例如,第3辆汽车在第8分钟末到达(即第9分钟之初),其服务开始时刻是第14分钟初,所以等待时间是5 min。

第10列所指的是该系统从前一次到达至这一次到达之间所经过的总空闲时间。空闲时间只是在到达状态为0时才出现。从本次到达时刻减去前一辆车的出发时刻即可求得空闲时间。第14辆汽车是在第49分钟末到达的,前一辆车结束服务时间是在第47分钟末,所以该系统空闲2 min。

15.5.3 工作记录表内容的说明

我们选择几辆汽车的到达项目加以讨论,以说明模拟工作记录表的应用。第1辆到达汽车出现在第3分钟末,由于系统此时是空闲的,所以这辆车可以加入系统,并立即进入洗刷设施。洗刷作业在第4分钟初开始,该车经过2 min的服务时间,因此,服务结束时刻是在第5分钟末。该车是在第4和第5分钟内进行洗刷的。第1辆到达的汽车不必排队等待,所以在等待线上的排队时间为0,并列入第9列。在第一个顾客到达之前,该服务系统空闲了3 min,因此3填入第10列中。

第4辆车到达是在第14分钟末,它所遇到的情况与第1辆车有所不同。当该顾客到达汽车洗车场时,第3辆车还在系统中,并将持续到第16分钟末。因此到达状态的项目内应记为1,这表明当第4辆车到达时,前一顾客还在系统内。但该顾客仍然加入系统,因为两个停车车位都可使用,须等到第17分钟初,才能开始服务。第四个顾客的等待时间是2 min(即第15分钟和第16分钟),将其填入第9列中。该顾客的服务时间是3 min,所以结束服务时刻是在第19分钟末。在这辆汽车到达前,系统没有空闲时间,记入第10列中。

第 8 辆车到达的情况是非常有趣的。该系统已经饱和，其状态为 3，第 8 辆车将不可能加入该系统，在第 6 列就表明了这一点。该系统处于饱和状态，其直接原因是由于第 5 辆车延长了洗刷时间和后续车到达间隔较小。这一情况就使得系统丧失了第 8 辆、第 9 辆和第 11 辆车。

15.5.4 模型参数的变化

每一个有希望采用的系统或有价值的解决方案都应进行模拟，并且应建立一套评价系统优缺点的方法。这种处理方法的细节，在反复运用中是要发生变化的，以使其能够适合某一具体的情况。此外，在分析系统时，通常还应包括有形费用、无形费用以及利润等。

一旦初始的模拟运行完成以后，接着应改变模型的作业特点，并重复进行模拟。在比尔洗车场的例子中，我们将进一步研究将停车车位数目由 2 个增加到 4 个的效果。

表 15 - 15 是一个模拟工作的记录表，它表示在修改后的作业条件下，有关前 20 名到达顾客的计算结果。表 15 - 11 中的顾客到达时刻、汽车型号和服务时间等仍作为输入数据再次使用。因此，第二次模拟是在与初始模拟有相同的输入或需求条件下进行的，这是蒙特卡罗模拟的主要优点：由于可以利用精确的随机数序列或输入的数值，我们就能够在相同的需求条件下，对两个或多个可供选择系统的方案作出评价。这样做可以消除其他因素所引起系统特性差异的可能性，而这些差异并不是分析人员规定要改变的。从而就可以利用较小规模的抽样，在满足给定的统计标准的条件下，确定两个方案的特殊差异。

表 15 - 15　等待线最大容量为 4 辆汽车的情况下，比尔洗车场 20 次汽车到达的模拟

时间单位：min

(1) 到达号	(2) 达到时刻	(3) 到达类型	(4) 服务时间	(5) 到达状态	(6) 是否加入系统	(7)* 开始服务时刻	(8)* 结束服务时刻	(9) 排队时间	(10) 空闲时间
1	3	Ⅰ	2	0	是	4	5	0	3
2	7	Ⅱ	6	0	是	8	13	0	2
3	8	Ⅰ	3	1	是	14	16	5	0
4	14	Ⅰ	3	1	是	17	19	2	0
5	18	Ⅱ	6	1	是	20	25	1	0
6	21	Ⅱ	4	1	是	26	29	4	0
7	23	Ⅱ	4	2	是	30	33	6	0
8	23	Ⅰ	2	4	是	34	35	10	0
9	24	Ⅰ	3	4	是	36	38	11	0

续 表

(1) 到达号	(2) 达到 时刻	(3) 到达 类型	(4) 服务 时间	(5) 到达 状态	(6) 是否加 入系统	(7) * 开始服 务时刻	(8) * 结束服 务时刻	(9) 排队 时间	(10) 空闲 时间
10	27	Ⅱ	5	5	否	39	43	11	0
11	28	Ⅰ	2	2	是	—	—		
12	35	Ⅰ	2	2	是	44	45	8	0
13	43	Ⅱ	4	4	是	46	49	2	0
14	49	Ⅰ	3	3	是	50	52	0	0
15	53	Ⅱ	4	4	是	54	57	0	1
16	54	Ⅰ	2	2	是	58	59	3	0
17	63	Ⅱ	4	4	是	64	67	0	4
18	66	Ⅱ	6	6	是	68	73	1	0
19	71	Ⅰ	3	3	是	74	76	2	0
20	72	Ⅱ	8	8	是	77	84	4	0

注：假定开始服务时刻是由该分钟初算起，结束服务在该分钟末。

在新的作业特点情况下，特别重要的事实是：前 20 名到达的顾客中仅仅失去一名顾客。在第 8 个和第 9 个顾客在初始模拟中都是要失去的，而现在可以进入该系统了。第 11 个顾客在第 28 分钟到达系统，此时第 6 个顾客正在接受服务，并且 4 个停车场车位都被另外 4 名顾客所占用，因此它仍不能进入该系统。同时，顾客的等待时间增加，前 20 个到达的顾客花费在排队等待上的总时间为 70 min，这将近比第一次模拟的排队总时间 37 min 多一倍。尽管有这些差异，在第 20 名顾客到达时，该系统所处的状态是相同的。

15.5.5 方案比较

在比尔洗车场的例子中，方案的评价是十分简单的，它只要确定由于增加了洗刷汽车的数目所得到的收益，能否补偿改变服务系统所支出的费用。

假定每辆汽车的洗刷费用为 3 美元，租用两个车位的停车场，其年租金换算为每营业小时的费用是 0.10 美元。通过减少服务设施的空闲时间可以增加洗刷汽车的数目，而洗刷这些车辆只需要直接开支，即包括水、洗浆剂和其他消耗品的费用。詹姆斯先生估计这些费用平均每辆汽车为 0.25 美元，因此，洗刷的汽车每增加一辆，就可提供 2.75 美元来补偿停车场的租金。只要收入超过他的支出就是有利的。增加停车车位的措施，可以增加每小时洗刷的汽车数近似为 0.8 辆，故每小时的期望的利润约为 2.10(0.8×2.75－0.10＝2.10)美元。因此，停车车位应该由两个增加到四个。尽管利润增加 2.10 美元的估计会有统计误差，但是这种增加幅度对于确保詹姆斯先生利润的增长是足够的。

排队时间的急剧增加，可能影响到顾客是否愿意参加排队的决定，这是一个潜在的困难，它可能使得上述结论无效。某些顾客在他们到达汽车洗刷场时，该系统已有三辆汽车正在等待，尽管这时还有一个停车车位可以使用，他们也可能决定到另一个洗车场去。顾客的这种行动会使增加停车车位的有利性降低。如果詹姆斯先生认为这种情况是可能的话，他就应该力图对这种影响进行定量的分析。但这是一项困难的工作。

15.6 模拟分析的例子

模拟方法可以而且已经应用于很多不同的场合。模拟对于排队问题的应用已经被证明是有效的，正如前面的例子所展示的一样。模拟还可以运用于库存控制、生产进度安排、仓库配置、装配线的平衡、有价证券的分析、经理人员训练以及其他情况。此外，模拟还可能有其他的用途。例如描述一个现有的系统或者设计一个新的系统。在本节中我们将就模拟的另外 3 个领域，列举一些简单的例子。

15.6.1 库存问题的模拟

某公司希望制定一个库存策略，即确定对某一特定产品的订货数量和订货点。这个公司面临两个方面的变化：需求量的变化以及订货与收到订货之间所需时间（交货期限）的变化。日需求量的概率分布如表 15－16 所示。交货期限的概率分布如表 15－17 所示，其变化范围为 1～6 天。每天每件商品的存储费用为 5 美元，脱销费用或丢失销售的费用为每件 30 美元，安排一次订货的费用为 80 美元。该公司希望制定一个库存策略，使得库存的存储费、脱销费用和订货费用的总和最小。

表 15－16 某一库存模拟需求量分布及相关联的随机数

需求量件数	概　率	随机数
8	0.13	01～13
9	0.19	14～32
10	0.34	33～66
11	0.23	67～89
12	0.11	89～00
	1.00	

表 15－17 某一库存模拟交货时间的概率分布和相关联的随机数

交货天数	概　率	随机数
1	0.14	01～14
2	0.34	15～48

续 表

交货天数	概 率	随机数
3	0.20	49～68
4	0.12	69～80
5	0.10	81～90
6	0.10	91～00
	总计 1.00	

解决这个问题的一个方法，是利用经济订货量和期望值的分析，来建立一个策略。建立策略的方法之一是采用模拟，通过对不同的订货量和订货点进行试验可以得到模拟的结果。为了产生我们所需要的模拟输入量，其随机数的分布列入表15-16和表15-17中。实际采用蒙特卡罗方法产生的随机变量的输入数值如表15-18的模拟工作记录表所示。

对于库存问题的分析，其模拟时间的推移是采用固定增量的方法。我们将时间每次前进一天而不是前进到下一个有关事件发生的时刻，后者是我们前面已经采用过的可变增量的方法。选择一天作为时间增量，必须假定订货总是安排在一天的终了，而新运到的货物的供应则在一天的开始。

公司已决定首先模拟这样一个策略，即每当库存下降到30件时就安排订货60件。在模拟开始时，库存是60件。在第三天末库存已经下降到订货点30件以下，因此要安排订货60件。随机数54用以确定这次订货的交货时间，由表15-17查得为3天。这次订货看作是3天以后收到，即在第六天末收到订货，它可以保证第七天的供应。因此，收到的60件订货是在第七天加入库存。在收到订货以前，库存已经用尽，并出现脱销，在第六天销售损失为2件。

这个模拟过程延续了40天。在这个经营周期内存储费是480.50美元，脱销费用是1 410美元，6次订货总共的订货费用为480美元，由此得出日平均费用为59.26美元。对于制定决策的目的来说，模拟持续时间有必要做相当的延长。

对于若干个不同的库存策略都应该重复进行模拟，并对各个方案的费用进行比较。根据第一次试验的有限结果，公司可能得出结论：脱销费用过高。于是将改变其初始策略，以减少脱销费用。为此可将订货点提高为30件以上。原来的订货点取为等于交货期限内的期望需求量。由表15-17和表15-18中给出的概率分布可算出每日期望需求量为10件，期望交货期限是3天，故在这段时间内的期望需求量是30件。提高订货点将增加库存的存储费，但这部分增加的费用将会由减少脱销费用所补偿。

表 15－18　某一库存问题的模拟工作记录

日期	收货	开始库存	随机数	需求量	终结库存	销售损失	随机数	交货日期
1		60	83	11	49			
2		49	70	11	38			
3		38	59	10	28		54	3
4		28	12	8	20			
5		20	92	12	8			
6		8	46	10	0	2		
7	60	60	54	10	50			
8		50	04	8	42			
9		42	51	10	32			
10		32	99	12	20		90	5
11		20	84	10	10			
12		10	81	11	0	1		
13		0	15	9	0	9		
14		0	36	0	0	10		
15		0	12	8	0	8		
16	60	60	54	10	50			
17		50	97	12	38			
18		38	00	12	26		71	4
19		26	49	10	16			
20		16	44	10	6			
21		6	13	8	0	2		
22		0	23	9	0	9		
23	60	60	45	10	50			
24		50	54	10	40			
25		40	24	9	31			
26		31	50	10	21		08	1
27		21	29	9	12			
28	60	72	61	10	62			
29		62	22	9	53			
30		53	47	10	43			
31		43	73	11	32			
32		32	18	9	23		51	3
33		23	22	9	14			
34		14	38	10	4			
35		4	34	10	0			
36	60	60	54	10	50			
37		50	91	12	38			
38		38	42	10	28		73	4
39		28	46	10	18			
40		18	30	9	9			
				961	47			

15.6.2 失效时间的模拟

哈特门公司研制了一种新的电子系统，并希望对新建立系统的无修理运行的平均持续时间做一个估计。此系统由 3 个标准装配而成。当这些部件中有一个失效时，该系统就失效并进行修理。

根据系统实际运行情况来做出这个估计，则费用太高，所需的时间太长。该公司已经收集了每一部件的各种运行记录，这些记录已用于确定部件失效时间的概率分布，如表 15－19 所示。此分布可用来直接计算新系统失效时间的概率分布。另一种方案是大量的模拟系统的故障，作为估计所研究的分布的一种方法。

表 15－19　三种电子部件的失效时间分布及相关的随机数分布

A 部件			B 部件			C 部件		
失效时间	概率	随机数	失效时间	概率	随机数	失效时间	概率	随机数
4 个月	0.10	01～10	2 个月	0.05	01～5	6 个月	0.20	01～20
5 个月	0.20	11～30	3 个月	0.10	06～15	7 个月	0.30	21～50
6 个月	0.30	31～60	4 个月	0.20	16～35	8 个月	0.25	51～75
7 个月	0.20	61～80	5 个月	0.30	36～65	9 个月	0.15	76～90
8 个月	0.15	81～95	6 个月	0.25	66～90	10 个月	0.10	91～00
9 个月	0.05	96～00	7 个月	0.10	91～00			

随机数的分布如表 15－19 所示，利用它可以产生出 3 种部件的失效时间。表 15－20 是模拟工作记录表，表示了 20 台该种电子装置运行间隔或失效时间的模拟。利用表 15－5 中的随机数，参照表 15－19 中给出的随机数分布，可以确定出三种部件(A、B、C)的失效时间。在第一种情况中，A、B、C 部件的失效时间分别求得为 6 个月、4 个月和 8 个月。B 部件的最早失效时间(4 个月)决定了整个系统的失效时间。根据对 20 个系统的模拟，得出系统平均失效时间是 4.6 个月。根据模拟结果的相对频率，模拟应进一步扩展使能提供一个更精确的运行时间估计。

表 15－20　某一电子系统失效时间的模拟工作记录表

系　统	随机数	A 部件失效月数	随机数	B 部件失效月数	随机数	C 部件失效月数	系统失效月数
1	33	6	24	4	52	8	4
2	50	6	72	6	85	9	6
3	13	5	19	4	79	9	4

续 表

系 统	随机数	A部件失效月数	随机数	B部件失效月数	随机数	C部件失效月数	系统失效月数
4	82	8	20	4	86	9	4
5	59	6	91	7	72	8	6
6	30	5	88	6	20	6	5
7	24	5	95	7	12	6	5
8	02	4	38	5	21	7	4
9	15	5	41	5	99	10	5
10	38	6	51	5	58	8	5
11	12	5	08	3	04	6	3
12	85	8	23	4	36	7	4
13	92	8	55	5	01	6	5
14	79	7	27	4	84	9	4
15	59	6	80	6	13	6	6
16	11	5	26	4	06	6	4
17	97	9	54	5	15	6	5
18	39	6	47	5	73	8	5
19	71	7	14	3	64	8	3
20	16	5	59	5	96	10	5

15.6.3 风险分析的模拟

蒙特卡罗模拟方法也常用于从事投资风险的分析，以下将举例说明如何用于分析推出新产品的投资问题。此项投资的盈利能力需视下列因素而定，即该项产品市场的预估、本公司可能获得的市场占有率、市场的成长率、生产成本、销售价格、产品的寿命以及所需要的设备成本等，但是以上这些因素都只能做不确定的估计。

通常所用的方法是对上述各项不确定因素作一个较好的单一估计，然后再计算获利能力，例如投资净现值或者报酬率等。用此种单一估计的方法分析投资效益有下面两个缺点：

(1) 用单一估计所获得的预期利润，可能与实际情形不相符。

(2) 无法测度投资的风险，即不能确定投资计划的获利或发生亏损的概率。

风险分析即是用于克服以上两个缺点的一种运算技巧。通常是先对每个不确定因素设定一个主观的概率分布，然后再运用蒙特卡罗法，将这个概率分布结合成以投资计划获利性为整体的一项概率分布。现在用一个简单的例子说明如下：

假定推出某种新产品需要投资固定设备 5 000 美元，有三种不确定的因素需要考虑，即售价、变动成本及销售量。此项产品的寿命只有一年。表 15－21 中含有上述不确定因素的各种情况及其各自可能发生的概率。为了便于预算，设定表中各因素在统计上是独立的，否则，将因概率的相依而需要对模拟过程加以修正。

表 15－21 风险分析中的不确定因素的概率分布

售价/美元	概　率	变动成本/美元	概　率	销售量/千	概　率
4 000	0.3	2 000	0.1	3 000	0.2
5 000	0.5	3 000	0.6	4 000	0.4
6 000	0.2	4 000	0.3	5 000	0.4

在这个简单的例子中，只有 $3\times3\times3=3^3=27$ 种可能的不同结果，运用**决策树(decision tree)**的分析方法也可以计算出各种结果和概率。但是在较复杂的实际问题中，可能含有较多的不确定因素，而每种因素又可能有 10 种或 20 种不同情况，整个问题即可能有上百万的可能结果。在此种情形下，将模拟的技巧应用于投资计划的平均获利能力及其风险的估计则非常有效，可以用实现各种不同利润水平的概率来计算。

为了能够依照表 15－21 中不确定因素的各种情况所列出的概率而产生随机数，像表 15－19 一样，可按各种结果可能出现的概率而设定若干相对应的随机数，如表 15－22 中所列。

表 15－22 对应于不确定因素的随机数

售价/美元	随机数	变动成本/美元	随机数	销售量/千	随机数
4 000	0～2	2 000	0	3 000	0～1
5 000	3～7	3 000	1～6	4 000	2～5
6 000	8～9	4 000	7～9	5 000	6～9

现在可以依照表 15－22 中所设定的与各不确定因素相对应的随机数，从随机数表 15－5 中选出一位的随机数，再依次决定售价、成本及销售量。这几个因素一旦决定，便可运用下式计算利润，即

$$利润=(售价-成本)\times销售量-固定费用$$

重复运用此式多次，即可产生许多种不同的利润情况，表 15－23 是重复模拟 25 次的结果。25 次的小型取样当然不足以完全正确地估计出平均盈利能力，如果将此种模拟过程设计成程序，而交由计算机去做，可以很容易地模拟几百次，甚至

上千次，即可获得较正确的结果。表 15－23 是这 25 次的模拟结果。为方便说明，下面的讨论都是基于表 15－23。

表 15－23 风险分析问题中模拟 25 次的结果

模拟次数	随机数	售价/美元	随机数	成本/美元	随机数	销售量/千	利润/美元
1	8	6 000	0	2 000	6	5	15 000
2	0	4 000	4	3 000	3	4	—1 000
3	6	5 000	2	3 000	2	4	3 000
4	1	4 000	4	3 000	0	3	—2 000
5	3	5 000	6	3 000	0	3	1 000
6	5	5 000	6	3 000	9	5	5 000
7	1	4 000	6	3 000	7	5	0
8	3	5 000	8	4 000	8	5	0
9	2	4 000	8	4 000	6	5	—5 000
10	1	4 000	6	3 000	1	3	—2 000
11	5	5 000	7	4 000	3	4	—1 000
12	9	6 000	9	4 000	6	5	5 000
13	4	5 000	9	4 000	7	5	0
14	7	5 000	2	3 000	6	5	5 000
15	9	6 000	5	3 000	3	4	7 000
16	0	4 000	5	3 000	0	3	—2 000
17	1	4 000	1	3 000	8	5	0
18	0	4 000	6	3 000	4	4	—1 000
19	8	6 000	8	4 000	6	5	5 000
20	9	6 000	2	3 000	4	4	7 000
21	0	4 000	7	4 000	7	5	—5 000
22	0	4 000	0	2 000	8	5	5 000
23	4	5 000	0	2 000	1	3	4 000
24	6	5 000	5	3 000	8	5	5 000
25	4	5 000	0	2 000	1	3	4 000
	平均						2 080

由表中的资料可以计算出，平均利润是 2 080 美元，将此项结果和简单的分析方法所获得的结果可以做一个比较。若用各不确定因素的最可能值为单一估计

值,则估计利润为

$$最可能利润 = (5\,000 - 3\,000) \times 4\,000 - 5\,000 = 3\,000(美元)$$

由此可知,单数估计的结果比模拟所得的预期利润相差很多。

在这个简单的例子里,可以将 27 种结果各自依概率加权而计算出预期利润,此项利润为 2 140 美元。由此可知,模拟 25 次所得的平均利润与计算得到的预期利润并不完全相同,但是预期利润的计算无法得到投资计划所附带的风险,然而模拟的 25 次结果中却清楚地显示出在何种情况下会发生亏损,例如在表 15 - 23 中,第二、第四、第九和第十等各次所模拟的结果。为了便于显示在风险分析中所获知的风险程度,可依表 15 - 23 中获利能力而排列层次,用图解来表示样本的累积概率密度函数,如图 15 - 3 所示。

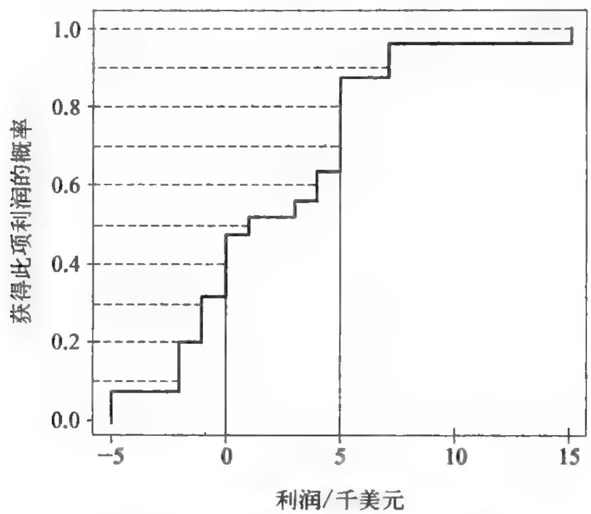


图 15 - 3　风险分析(样本累计概率函数)

由图 15 - 3 中可以看出,有 68%的机会可以获得利润(即有 32%的机会将蒙受损失),有 36%的机会可以获得利润 5 000 或 5 000 美元以上,获利最多为 15 000 美元,其可能性为 4 %。若模拟的次数非常多,图 15 - 3 中的阶梯线段将逐渐趋近光滑,但并不能成为连续的曲线,因为各种不同的结果是间断的。

15.7　模拟的优点和局限性

在问题分析中广泛地使用模拟,反映了模拟方法具有一系列的优点,其中一些

优点前面已提到过，在本节内容中我们做进一步的补充。

模拟可以用来研究大型复杂的项目，这些项目包含着随机性的多种因素，这是使用模拟的主要原因。此外，往往由于选择不出有效的分析方法，而存在这样一个问题：是采用模拟还是不采用模拟去研究它。唯一的另一种方法，就是用实际的系统进行试验，不过这往往是不切合实际的或者说是不可能的。模拟具有强大的功能，其主要原因之一，就是我们可以把更详细、更接近实际的内容纳入模拟模型，这一点是可求解的分析模型通常办不到的。

模拟为分析人员和管理人员提供了一个“实验室”，它使得分析人员和管理人员在相当短的时间内，能够预见所研究的系统在一定运行条件下的特性。它还使得他们能够研究单个变量或参数的变化对现有系统的影响。在比尔汽车洗车场模拟的第二个方案中，我们可以了解到由于停车车位从 2 个变到 4 个时该系统特性的任何差异，而其他部分仍保持不变。这一点在实际系统中是不可能的，因为在同一时间内会发生各种变化，因此要把一个给定的改变措施的效果分离出来是困难的，或者说是不可能的。此外，模拟允许分析人惯用某一特定变量和参数的相同数值，去充分试验，这在实际系统中也是不可能的。

模拟的一个更大的优点就是其基本方法简单，易于领会。模拟得到的结果易于为管理人员理解，而是否采纳和应用这些结果是由他们决定的。

遗憾的是，模拟除了具有通常有关模型构造的缺点外，还有某些其他的缺点，所以当我们考虑使用模拟时，必须权衡这些缺点。切记模拟模型是运行的，而不是被求解的。根据精心设计的模拟研究所选择的系统方案，只是已经考虑过的各种方案中的最佳方案。然而事实上不可能对所有可能的系统都考虑一遍，所以无法知道所选择的方案是否为最有效。与数学求解相反，模拟模型的一个根本的缺点，就是不能保证得到模型的最优解。因此，分析人员必须对那些选择方案的人的能力具有相当的信心。他们的经验和洞察力，应使得所选择的方案即使不是最优的，也应该是好的系统。然而，系统复杂性增加时，这个任务就变得特别困难。

模拟的第二个主要缺点是对系统特性估计的不精确性。这些估计存在有统计误差。即使在设计得最好的试验中，要确定哪一个研究方案为最佳方案，也是困难的或是不可能的。其中的部分困难是由于还没有完全研究出一种恰当和易于应用的统计方法，以分析和比较模拟得出的结果。

模拟处理方法的另一个重要缺点，是研制模型和进行模拟的费用高。为了分析复杂的系统需要研究计算机程序，其费用可能是非常大的。进行模拟所需要的计算机时间的费用支出也会限制所使用的模型的复杂程度。考虑到费用问题，就可能迫使构造模拟模型的人员采取与分析模型中同样的方式，对模拟模型的假设条件进行简化。

模拟所具有的这些缺点，意味着分析人员应该把模拟看作是分析方法的备用，而不能用模拟代替分析方法。如果有一个分析方法适合于采用，就应该采用它，而不是去做模拟。分析人员必须防止出现把模拟当作分析工具，而过分依赖于模拟的倾向。只有经过仔细研究，确定不能采用分析方法时，才应使用模拟。

附表

附表 1　t 检验临界值表

$P(|y|>t_{\alpha})=\alpha$

f \ α	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.05	0.02	0.01	0.001	α / f
1	0.158	0.325	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619	1
2	0.142	0.289	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.598	2
3	0.137	0.277	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924	3
4	0.134	0.271	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610	4
5	0.132	0.267	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.859	5
6	0.131	0.265	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959	6
7	0.130	0.263	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.405	7
8	0.130	0.262	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041	8
9	0.129	0.261	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781	9
10	0.129	0.260	0.397	0.542	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587	10
11	0.129	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437	11
12	0.128	0.259	0.395	0.539	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318	12
13	0.128	0.259	0.394	0.538	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221	13
14	0.128	0.258	0.393	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140	14
15	0.128	0.258	0.393	0.536	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073	15

续 表

f \ α	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.05	0.02	0.01	0.001	α / f
16	0.128	0.258	0.392	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015	16
17	0.128	0.257	0.392	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965	17
18	0.127	0.257	0.392	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922	18
19	0.127	0.257	0.391	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883	19
20	0.127	0.257	0.391	0.533	0.687	0.860	1.064	1.325	1.725	2.086	2.538	2.845	3.850	20
21	0.127	0.257	0.391	0.532	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819	21
22	0.127	0.256	0.390	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792	22
23	0.127	0.256	0.390	0.532	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.767	23
24	0.127	0.256	0.390	0.531	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745	24
25	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725	25
26	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707	26
27	0.127	0.256	0.389	0.531	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690	27
28	0.127	0.256	0.389	0.530	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674	28
29	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659	29
30	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646	30
40	0.126	0.254	0.388	0.529	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551	40
60	0.126	0.254	0.387	0.527	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.760	3.460	60
120	0.126	0.254	0.386	0.526	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373	120
∞	0.126	0.253	0.385	0.524	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291	∞

附注：α 为显著性水平；f 为自由度。

附表 2　*F* 检验临界值表

$\alpha = 0.01$

u_2 \ u_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4 052	5 000	5 403	5 625	5 764	5 859	5 928	5 982	6 022	6 056	6 106	6 157	6 209	6 235	6 261	6 287	6 313	6 339	6 366
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.6	99.5
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.7	26.6	26.5	26.4	26.3	26.2	26.1
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.9	13.8	13.7	13.7	13.6	13.5
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.70	7.31	7.23	7.14	7.06	6.97	6.88
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.3	8.65	7.52	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	5.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	5.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.90	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.20	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

续 表

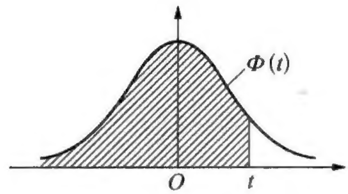
$\alpha = 0.05$

u_2 \ u_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	245	248	249	250	251	252	253	254
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	4.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	0.37
16	4.49	3.63	3.24	3.10	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	0.23	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.51	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

附注：u_1 为第 1 自由度；u_2 为第 2 自由度。

附表 3 标准正态分布表

$$\Phi(t)=\int_{-\infty}^{t}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$$



t	0	1	2	3	4	5	6	7	8	9
−3.0	0.001 3	0.001 0	0.000 7	0.000 5	0.000 3	0.000 2	0.000 2	0.000 1	0.000 1	0.000 0
−2.9	0.001 9	0.001 8	0.001 7	0.001 7	0.001 6	0.001 6	0.001 5	0.001 5	0.001 4	0.001 4
−2.8	0.002 6	0.002 5	0.002 4	0.002 3	0.002 3	0.002 2	0.002 1	0.002 1	0.002 0	0.001 9
−2.7	0.003 5	0.003 4	0.003 3	0.003 2	0.003 1	0.003 0	0.002 9	0.002 8	0.002 7	0.002 6
−2.6	0.004 7	0.004 5	0.004 4	0.004 3	0.004 1	0.004 0	0.003 9	0.003 8	0.003 7	0.003 6
−2.5	0.000 2	0.006 0	0.005 9	0.005 7	0.005 5	0.005 4	0.005 2	0.005 1	0.004 9	0.004 8
−2.4	0.008 2	0.008 0	0.007 8	0.007 5	0.007 3	0.007 1	0.006 9	0.006 8	0.006 6	0.006 4
−2.3	0.010 7	0.010 4	0.010 2	0.009 9	0.009 6	0.009 4	0.009 1	0.008 9	0.008 7	0.008 4
−2.2	0.013 9	0.013 6	0.013 2	0.012 9	0.012 6	0.012 2	0.011 9	0.011 6	0.011 3	0.011 0
−2.1	0.017 9	0.017 4	0.017 0	0.016 6	0.016 2	0.015 8	0.015 4	0.015 0	0.014 6	0.014 3
−2.0	0.022 8	0.022 2	0.021 7	0.021 2	0.020 7	0.020 2	0.019 7	0.019 2	0.018 8	0.018 3
−1.9	0.028 7	0.023 1	0.027 4	0.026 8	0.026 2	0.025 5	0.025 0	0.024 4	0.023 8	0.023 3
−1.8	0.035 9	0.035 2	0.034 4	0.033 6	0.032 9	0.032 2	0.031 4	0.030 7	0.030 0	0.029 4
−1.7	0.044 6	0.043 6	0.042 7	0.041 8	0.040 9	0.040 1	0.039 2	0.038 4	0.037 5	0.036 7
−1.6	0.054 8	0.053 7	0.052 6	0.051 6	0.050 5	0.049 5	0.048 5	0.047 5	0.046 5	0.045 5
−1.5	0.066 8	0.065 5	0.064 3	0.063 0	0.061 8	0.060 6	0.059 4	0.058 2	0.057 0	0.055 9
−1.4	0.080 8	0.079 3	0.077 8	0.076 4	0.074 9	0.073 5	0.072 2	0.070 8	0.069 4	0.068 1
−1.3	0.096 8	0.095 1	0.093 4	0.091 8	0.090 1	0.088 5	0.086 9	0.085 3	0.083 8	0.082 3
−1.2	0.115 1	0.113 1	0.111 2	0.109 3	0.107 5	0.105 6	0.103 8	0.102 0	0.100 3	0.093 5
−1.1	0.135 7	0.133 5	0.131 4	0.129 2	0.127 0	0.125 1	0.123 0	0.121 0	0.119 0	0.117 0
−1.0	0.158 7	0.156 2	0.153 9	0.151 5	0.149 2	0.146 9	0.144 0	0.142 3	0.140 1	0.137 9
−0.9	0.184 1	0.181 4	0.178 8	0.176 2	0.173 6	0.171 1	0.168 5	0.166 0	0.163 5	0.161 1
−0.8	0.211 9	0.209 0	0.206 1	0.203 3	0.200 5	0.197 7	0.194 9	0.192 2	0.189 4	0.186 7
−0.7	0.242 0	0.238 9	0.235 8	0.232 7	0.229 7	0.226 6	0.223 6	0.220 6	0.217 7	0.214 8
−0.6	0.274 3	0.270 9	0.267 6	0.264 3	0.261 1	0.257 8	0.254 6	0.251 4	0.248 3	0.245 1
−0.5	0.308 5	0.305 0	0.301 5	0.298 1	0.294 6	0.291 2	0.287 7	0.284 3	0.281 0	0.277 6

续　表

t	0	1	2	3	4	5	6	7	8	9
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.9711	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.9823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8641	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9430	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9648	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9700	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9762	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9825	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9874	0.9848	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9006	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9984	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000

参考文献

[1] 高百宁.经济预测与决策[M].上海：上海财经大学出版社，2009.
[2] 徐国祥.统计预测与决策[M].上海：上海财经大学出版社，2010.
[3] 贾俊平，何晓群，金勇进.统计学[M].北京：中国人民大学出版社，2009.
[4] 暴奉贤，陈宏立.经济预测与决策方法[M].广州：暨南大学出版社，2004.
[5] 丁以中.管理科学——运用 Spreadsheet 建模和求解[M].北京：清华大学出版社，2003.
[6] 张照贵.管理决策模型、方法与应用[M].成都：西南财经大学出版社，2008.
[7] 郎茂祥.预测理论与方法[M].北京：清华大学出版社，北京交通大学出版社，2011.
[8] 陈友玲.市场调查、预测与决策[M].北京：机械工业出版社，2009.
[9] 简明，胡玉立.市场预测与管理决策(第四版)[M].北京：中国人民大学出版社，2009.
[10] 张健，王晖，陈元凤.管理决策——模型与应用[M].北京：机械工业出版社，2010.
[11] 易丹辉.数据分析与 Eviews 应用[M].北京：中国统计出版社，2002.
[12] 刘思峰，党耀国.预测方法与技术[M].北京：高等教育出版社，2005.
[13] 宁宣熙，刘思峰.管理预测与决策方法[M].北京：科学出版社，2003.
[14] 李锋编.经济预测与决策实验教程[M].广州：华南理工大学出版社，2008.
[15] 蔡美德，徐剑虹.管理决策分析[M].广州：华南理工大学出版社，1992.
[16] 王可定，周献中.运筹决策理论方法新编[M].北京：清华大学出版社，2010.